



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 105983**

**TO: Susan Ungar**  
**Location: CM1/8E12/8B05**  
**Art Unit: 1642**  
**Wednesday, October 15, 2003**

**Case Serial Number: 09/854124**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**CM1-6B02**  
**Phone: 305-9203**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Ungar,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

**THIS PAGE BLANK (USPTO)**

Hart, Edward

105983

**From:** Ungar, Susan  
**Sent:** Monday, October 13, 2003 2:23 PM  
**To:** Chan, Christina  
**Cc:** Hart, Edward  
**Subject:** Rush sequence search for 09/854,124

Hi

I need a rush sequence search for 09/854,124, SEQ ID NOS 5, 6, 7 - these are polynucleotide fragments.

Ed Hart has agreed to submit this search for me, please send authorization directly to him.

Thanks  
Susan Ungar  
1642  
703-305-2181  
CM1-8B05

10/14/03

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

ABSSP4  
3-NA

10/15/03

**THIS PAGE BLANK (USPTO)**



Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpn and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapn and .rapn

**Because they contain data that is confidential, the results of Pending database searches should not be left in the case .**



# STIC SEARCH RESULTS

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art found, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art not found:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library, CM1 - Circ. Des



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 13:00:56 ; Search time 1477 Seconds  
(without alignments)  
8309.335 Million cell updates/sec

Title: US-09-854-124-6

Perfect score: 300

Sequence: 1 aattccgtgtgtgtgtgtga.....cgaaaggagaacattttac 300

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rtd.\*
- 36: em.htg\_nam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	284.8	94.9	858	6	BD156103	BD156103 Primer fo
2	284.8	94.9	858	9	AK001275	AK001275 Homo sapi
3	284.8	94.9	1174	9	AF151880	AF151880 Homo sapi
4	284.8	94.9	1180	9	BC010164	BC010164 Homo sapi
5	284.8	94.9	1188	9	AF201939	AF201939 Homo sapi
6	284.8	94.9	1203	9	BC003186	BC003186 Homo sapi
7	284.8	94.9	3097	9	AK091519	AK091519 Homo sapi
8	283	94.3	667	6	BD147616	BD147616 Primer fo
9	227.6	75.9	116792	9	HS1100H13	BD135419 Human DNA
10	223.4	74.5	454	6	AX336730	AX336730 Sequence
11	167.4	55.8	933	5	AB097169	AB097169 xenopus l
12	125.2	41.7	160356	9	AC009068	AC009068 Homo sapi
13	125.2	41.7	185321	9	AC123908	AC123908 Homo sapi
14	120	40.0	245210	2	AC137771	AC137771 Homo sapi
15	97	32.3	207307	2	AC117085	AC117085 Rattus no
16	97	32.3	228373	2	AC118833	AC118833 Rattus no
17	93.8	31.3	188292	2	AC103360	AC103360 Mus muscu
18	85.2	28.4	825	3	AY071234	AY071234 Drosophill
19	85.2	28.4	72722	3	AC004373	AC004373 Drosophill
20	85.2	28.4	132666	3	AC005270	AC005270 Drosophill
21	85.2	28.4	165158	2	AC141687	AC141687 Apis mell
22	85.2	28.4	182897	3	AC092232	AC092232 Drosophill
23	85.2	28.4	218794	2	AC018084	AC018084 Drosophill
24	85.2	28.4	292393	3	AE003576	AE003576 Drosophill
25	59.2	19.7	33103	3	CEF31C3	Z92784 Caenorhabdi
26	55.4	18.5	4593	3	LEINACETYL	M96635 Leishmania
27	55.4	18.5	110000	2	LMFLCHR36_16	Continuation (17 o
28	51	17.0	245210	2	AC137771	AC137771 Homo sapi
29	47.6	15.9	139399	8	AP002865	AP002865 Oryza sat
30	47.6	15.9	182756	8	AC007789	AC007789 Oryza sat
31	47.2	15.7	37949	8	SPBC725	AL034352 S.pombe c
32	40.6	13.5	190459	2	AP005419	AP005419 Oryza sat
33	39.4	13.1	21024	8	AP002055	AP002055 Arabidops
34	39.4	13.1	105768	8	AC069474	AC069474 Arabidops
35	38.6	12.9	82804	3	BMMIF1	AL606837 Brugia ma
36	38.4	12.8	310550	1	SC0939113	AL939113 Streptomy
37	38	12.7	10596	1	AE004746	AE004746 Pseudomon
38	38	12.7	88223	2	AC084755	AC084755 Homo sapi
39	38	12.7	241891	2	AC132675	AC132675 Rattus no
40	38	12.7	259329	2	AC109570	AC109570 Rattus no
41	37.2	12.4	298450	1	SC0939107	AL939107 Streptomy
42	37	12.3	228940	2	AC126890	AC126890 Rattus no
43	36.8	12.3	7954	1	TAQ19223	Y19223 Thermus aqu
44	36.8	12.3	10869	1	AE005069	AE005069 Halobacte
45	36.4	12.1	3008	6	AX683109	AX683109 Sequence

ALIGNMENTS

RESULT 1  
BD156103  
LOCUS BD156103 858 bp DNA linear PAT 17-JAN-2003  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION BD156103  
VERSION BD156103.1 GI:27861861  
KEYWORDS JP 2002191363-A/10946.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 858)  
AUTHORS Ota.T., Isogai.T., Nishikawa.T., Hayashi.K., Saito.K., Yanamoto.J.,  
Ishii.S., Sugiyama.T., Wakamatsu.A., Nagai.K. and Otsuki.T.  
TITLE Primer for synthesizing full-length cDNA and use thereof

Patent: JP 2002191363-A 10946 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/10946  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OPA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC  
 C12N15/09.C07K14/47.C07K16/18.C12N1/15.C12N1/19.C12N1/21.C12N5/ PC  
 10, C12P21/02.C12Q1/68/C12P21/08.G06F17/30.C12N15/00.C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 Location/Qualifiers  
 FT CDS 227 a 214 c 222 g 195 t  
 (37)..(591).  
 FEATURES  
 source  
 1..858  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 227 a 214 c 222 g 195 t  
 ORIGIN  
 Query Match 94.9%; Score 284.8; DB 6; Length 858;  
 Best Local Similarity 99.3%; Pred. No. 2e-64;  
 Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 13 GTCGCTGAGGCTCGGCTCGAGCTCGCGCGCCATCGAGCTGCCGAGTCCGAATTCCT 72  
 Db 3 GACCGTGGAGCTCGGCGCTCGAGCTCGCGCGCCATCGAGCTGCCGAGTCCGAATTCCT 62  
 QY 73 GCCCGAGAAGGAGTGTACCATATCCCAACTTCAGTCTGGACAAAGATCACTCAT 132  
 Db 63 GCCCGAGAAGGAGTGTACCATATCCCAACTTCAGTCTGGACAAAGATCACTCAT 122  
 QY 133 CGGGGGGACCTGGGGCCCTTTTAAACCTGGTTTACCCTGGTGAAGTGCCTCGGCTGGC 192  
 Db 123 CGGGGGGACCTGGGGCCCTTTTAAACCTGGTTTACCCTGGTGAAGTGCCTCGGCTGGC 182  
 QY 193 GATTAACTGGAACAAAGACAGAAATGCGCTGCTCCCTCCAGAGTGGATGATGATAGA 252  
 Db 183 GATTAACTGGAACAAAGACAGAAATGCGCTGCTCCCTCCAGAGTGGATGATGATAGA 242  
 QY 253 AAAGTTGGAGAAGATGAGGGATCATGAACGAAAGAGAAACCTTTTAC 300  
 Db 243 AAAGTTGGAGAAGATGAGGGATCATGAACGAAAGAGAAACCTTTTAC 290  
 RESULT 2  
 AK001275 858 bp mRNA linear PRI 01-AUG-2002  
 LOCUS  
 Homo sapiens cDNA FLJ10413 fis, clone NT2RP1000063.  
 AK001275  
 DEFINITION  
 Homo sapiens cDNA FLJ10413 fis, clone NT2RP1000063.  
 AK001275  
 VERSION  
 oligo capping; fis (full insert sequence).  
 KEYWORDS  
 Homo sapiens (human)  
 SOURCE  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 ISOGAI,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,  
 Wagatsuma,M., Hosoliri,T., Kaku,Y., Kodaira,H., Kondo,H.,  
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,  
 Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 858)  
 ISOGAI,T. and OTSUKI,T.  
 TITLE  
 NEDO human cDNA sequencing project  
 JOURNAL  
 REFERENCE  
 1 (bases 1 to 858)  
 ISOGAI,T. and OTSUKI,T.  
 AUTHORS

TITLE Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics

JOURNAL Genome Res. 10 (5), 703-713 (2000)

MEDLINE 20272150

PUBMED 10810093

REFERENCE 2 (bases 1 to 1174)

AUTHORS Lin, W.-C.

TITLE Direct Submission

JOURNAL Submitted (17-MAY-1999) Institute of Biomedical Sciences, Academia Sinica, No. 128, Sec. II, Academia Road, Taipei 115, Taiwan

FEATURES

source

1..1174

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="16"

77..634

/codon\_start=1

/product="CGI-122 protein"

/protein\_id="AAD34117.1"

/db\_xref="GI:4929713"

/translation="MDAAVEFLAEKELVTIIPNLSLDKIYLGDLGPFNPGLPVEV PLWLAINLKQKCKLLPEWMDVEKLEKMRDHERKEETFTPMSPYYMELTKLLNH ASDNIPKADEITLVKMDWTRIAKLRSADSFVRQEAHAKLDNLTLMEINTSGTFL TQALNHMYKLRITNLPLESTQSDP"

BASE COUNT 300 a 299 c 307 g 268 t

ORIGIN

Query Match 94.9%; Score 284.8; DB 9; Length 1174;

Best Local Similarity 99.3%; Pred. No. 2.1e-64;

Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 GTCGTGAGGCTCTGGCCTGACGCTGCGCCGCCATGACGCTGCGGAGTGAATTCCT 72

DB 43 GACCGTGAGGCTCTGGCCTGACGCTGCGCCGCCATGACGCTGCGGAGTGAATTCCT 102

QY 73 CCGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 132

DB 103 CCGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 162

QY 133 CGGGGGGACCTGGGGCCTTTTAAACCCCTGTTTACCCGTGGAAGTCCCTCTGGCTGCGC 192

DB 163 CGGGGGGACCTGGGGCCTTTTAAACCCCTGTTTACCCGTGGAAGTCCCTCTGGCTGCGC 222

QY 193 GATTAACCTGAAACAAAGACAGAAATGTGCCTGCTCCCTCCAGAGTGGATGGATGTAGA 252

DB 223 GATTAACCTGAAACAAAGACAGAAATGTGCCTGCTCCCTCCAGAGTGGATGGATGTAGA 282

QY 253 AAAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 300

DB 283 AAAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 330

RESULT 4

BC010164

LOCUS

DEFINITION Homo sapiens, HSPC037 protein, clone MGC:19836 IMAGE:4098007, mRNA, complete cds.

ACCESSION BC010164

VERSION BC010164.1 GI:14603431

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1180)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1180)

Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (02-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: [http://www.nisc.nih.gov/nisc\\_mgc@nigr.nih.gov](http://www.nisc.nih.gov/nisc_mgc@nigr.nih.gov)

Contact: Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 27 Row: 1 Column: 5

This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.

FEATURES

source

1..1180

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="LocusID:51659"

/db\_xref="taxon:9606"

/clone="MGC:19836 IMAGE:4098007"

/tissue\_type="Brain, neuroblastoma"

/clone\_lib="NIH\_MGC\_19"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

68..625

/codon\_start=1

/product="HSPC037 protein"

/protein\_id="AAH10164.1"

/db\_xref="GI:14603432"

/translation="MDAAVEFLAEKELVTIIPNLSLDKIYLGDLGPFNPGLPVEV PLWLAINLKQKCKLLPEWMDVEKLEKMRDHERKEETFTPMSPYYMELTKLLNH ASDNIPKADEITLVKMDWTRIAKLRSADSFVRQEAHAKLDNLTLMEINTSGTFL TQALNHMYKLRITNLPLESTQSDP"

BASE COUNT 316 a 292 c 307 g 265 t

ORIGIN

Query Match 94.9%; Score 284.8; DB 9; Length 1180;

Best Local Similarity 99.3%; Pred. No. 2.1e-64;

Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 GTCGTGAGGCTCTGGCCTGACGCTGCGCCGCCATGACGCTGCGGAGTGAATTCCT 72

DB 34 GACCGTGAGGCTCTGGCCTGACGCTGCGCCGCCATGACGCTGCGGAGTGAATTCCT 93

QY 73 CCGCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 132

DB 94 CCGCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 153

QY 133 CGGGGGGACCTGGGGCCTTTTAAACCCCTGTTTACCCGTGGAAGTCCCTCTGGCTGCGC 192

DB 154 CGGGGGGACCTGGGGCCTTTTAAACCCCTGTTTACCCGTGGAAGTCCCTCTGGCTGCGC 213

QY 193 GATTAACCTGAAACAAAGACAGAAATGTGCCTGCTCCCTCCAGAGTGGATGGATGTAGA 252

DB 214 GATTAACCTGAAACAAAGACAGAAATGTGCCTGCTCCCTCCAGAGTGGATGGATGTAGA 273

QY 253 AAAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 300

DB 274 AAAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 321

RESULT 5

३.

QY 193 GATTAACTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGAGTGTAGA 252  
 |||||  
 Db 231 GATTAACTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGAGTGTAGA 290  
 |||||  
 QY 253 AAAGTTGGAGAGATGAGGGATCATGAACGAAGGAAGAAACTTTTAC 300  
 |||||  
 Db 291 AAAGTTGGAGAGATGAGGGATCATGAACGAAGGAAGAAACTTTTAC 338  
 |||||

RESULT 7  
 AK091519  
 LOCUS 3097 bp mRNA linear PRI 15-JUL-2002  
 DEFINITION Homo sapiens CDNA FLJ34200 fls, clone FCBF3019663.  
 AK091519  
 VERSION AK091519.1 GI:21749909  
 KEYWORDS oligo capping; fls (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,  
 Ono,Y., Hotuta,T., Hirao,K., S., Murakawa,K., Takiguchi,S.,  
 Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,  
 Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,  
 Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
 Kawai-Hio,Y., Sekine,K., Nishikawa,T., Kimura,K., Matsuo,K.,  
 Nakamura,Y., Sakino,M., Kikuchi,H., Kanda,K., Wagatsuma,M.,  
 Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,  
 Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
 NEDO human cDNA sequencing project

Unpublished  
 2 (bases 1 to 3097)  
 REFERENCE  
 AUTHORS Isogai,T. and Yamamoto,J.

TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7  
 Isogai,T. and Yamamoto,J.  
 Kuzusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.

FEATURES  
 source  
 1. 3097  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="FCBF3019663"  
 /tissue\_type="brain"  
 /clone\_lib="FCBF3"  
 /dev\_stage="fetal"  
 /note="cloning vector: pME18SFL3"  
 BASE COUNT 726 a 703 c 786 g 882 t  
 ORIGIN

Query Match 94.9%; Score 284.8; DB 9; Length 3097;  
 Best Local Similarity 99.3%; Pred. No. 2.2e-64;  
 Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 GTCGGTGAGGCTCTGGCGCTGCGCGCGCCATGACGCTGCCGAGGTCGAATTCCT 72  
 |||||  
 Db 1 GACCGTGAGGCTCTGGCGCTGCGCGCGCCATGACGCTGCCGAGGTCGAATTCCT 60  
 |||||  
 QY 73 CGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGCTGGACAAGATCTACCTCAT 132  
 |||||  
 Db 61 CGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGCTGGACAAGATCTACCTCAT 120  
 |||||  
 QY 133 CGGGGGGACCTGGGGCCCTTTAACCCCTGTTTACCCCGTGAAGTGGCCCTGTGGCTGGC 192  
 |||||

Db 121 CGGGGGGACCTGGGGCCCTTTTAAACCTGGTTTACCCGTGGAGTGCCTCTGGCTGGC 180  
 |||||  
 QY 193 GATTAACTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGAGTGTAGA 252  
 |||||  
 Db 181 GATTAACTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGAGTGTAGA 240  
 |||||  
 QY 253 AAAGTTGGAGAGATGAGGGATCATGAACGAAGGAAGAAACTTTTAC 300  
 |||||  
 Db 241 AAAGTTGGAGAGATGAGGGATCATGAACGAAGGAAGAAACTTTTAC 288  
 |||||

RESULT 8  
 BD147616  
 LOCUS 667 bp DNA linear PAT 17-JAN-2003  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
 AK091519  
 ACCESSION BD147616  
 VERSION BD147616.1 GI:27853374  
 KEYWORDS JP 2002191363-A/2459.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
 Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
 Primer for synthesizing full-length cDNA and use thereof  
 Patent: JP 2002191363-A 2459 09-JUL-2002;  
 JOURNAL HELIX RESEARCH INSTITUTE

COMMENT OS Homo sapiens (human)  
 PN JP 2002191363-A/2459  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 200280990  
 PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,  
 PI KEIICHI NAGAI,TETSUJI OTSUKI  
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC  
 10,  
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key

FT source  
 1. 667  
 /organism="Homo sapiens"  
 /Location/Qualifiers  
 1. 667  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 171 a 176 c 177 g 137 t 6 others  
 ORIGIN

Query Match 94.3%; Score 283; DB 6; Length 667;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-64;  
 Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 TGAGGCTCTGGCGCTGCGCGCGCCATGACGCTGCCGAGGTCGAATTCCTCGCGG 77  
 |||||  
 Db 7 TGAGGCTCTGGCGCTGCGCGCGCCATGACGCTGCCGAGGTCGAATTCCTCGCGG 66  
 |||||  
 QY 78 AGAAGAGCTGTTTACCATTATCCCAACTTCAGTCTGGACAAGATCTACCTCATCGGGG 137  
 |||||  
 Db 67 AGAAGAGCTGTTTACCATTATCCCAACTTCAGTCTGGACAAGATCTACCTCATCGGGG 126  
 |||||

QY 138 GGGACCTGGGCGCTTTTAAACCTGGTTTACCCGTGGAAGTGCCTGTGGCTGGCGATTA 197  
 |||||  
 Db 127 GGGACCTGGGCGCTTTTAAACCTGGTTTACCCGTGGAAGTGCCTGTGGCTGGCGATTA 186  
 |||||

QY 198 ACCTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGGATGGATGTAGAAAGT 257  
 |||||  
 Db 187 ACCTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGGATGGATGTAGAAAGT 246  
 |||||





misc_feature	/note="L2 repeat: matches 2657. .2701 of consensus" complement(12943. .13346)
polyA_signal	/note="match: STS: Em:G24573"
polyA_site	/gene="dJ1100H13.1" 13349
repeat_region	/gene="dJ1100H13.1" 14509. .14637
gene	/note="MER5A repeat: matches 54. .182 of consensus" 15718. .22954
mRNA	/gene="dJ1100H13.2" join(<15718. .15865,19926. .20186) /gene="dJ1100H13.2" /product="dJ1100H13.2.2 (putative novel protein (isoform 2))"
	/note="match: ESTs: Em:AA776833 Em:AA972215" /evidence=not_experimental join(<15797. .15865,20520. .20634,22592. .22954) /gene="dJ1100H13.2" /product="dJ1100H13.2.1 (putative novel protein (isoform 1))"
	/note="match: ESTs: Em:AI990489 Em:AI656415 Em:AA974242" /evidence=not_experimental complement(15882. .16370) /note="match: STS: Em:HS927M24S" 16047. .16178 /note="4 copies 33 mer 91% conserved"
misc_feature	/note="MER91A repeat: matches 2. .115 of consensus" 1620. .16889
repeat_region	/note="AluJb repeat: matches 2. .289 of consensus" 17303. .17608
repeat_region	/note="AluSg repeat: matches 1. .306 of consensus" 17634. .17809
repeat_region	/note="MIR repeat: matches 79. .262 of consensus" 17825. .18011
repeat_region	/note="MIR repeat: matches 36. .238 of consensus" 19117. .19233
repeat_region	/note="MIR repeat: matches 126. .252 of consensus" 19275. .19565
repeat_region	/note="AluSc repeat: matches 1. .292 of consensus" 19584. .19698
polyA_site	/note="MIR repeat: matches 132. .261 of consensus" 20186
mRNA	/gene="dJ1100H13.2" join(<20613. .20748,22592. .22954) /gene="dJ1100H13.2" /product="dJ1100H13.2.3 (putative novel protein (isoform 3))"
	/note="match: ESTs: Em:AI632573" /evidence=not_experimental 20940. .21205
repeat_region	/note="L2 repeat: matches 2403. .2748 of consensus" 21246. .21354
repeat_region	/note="MIR repeat: matches 127. .244 of consensus" 21421. .21616
repeat_region	/note="L2 repeat: matches 1610. .1811 of consensus" 21977. .22107
repeat_region	/note="MER45C repeat: matches 1. .907 of consensus" 22215. .22528
	/note="AluSx repeat: matches 1. .311 of consensus"
Query Match	75.9%; Score 227.6; DB 9; Length 116792;
Best Local Similarity	89.2%; Pred. No. 3.6e-49;
Matches 257; Conservative	0; Mismatches 29; Indels 2; Gaps 1;
QY	13 GTCGGTAGGCTGTGGCCCTGCAGCTCGCGCCCGCATGGACGCTGCCGAGGTCGATTCCT 72
Db	50001 GACTGTAGGCCGAGCTGGCCACTCGCGCCCGCATGGAGCTGGCCGAGTCGAGTTCT 49942
QY	73 CGCCGAGAGGAGCTGGTTACCATATATCCCAACTTCAGTCGGACACAGATCTACCTCAT 132
Db	49941 GGCTGAGAGGAGCTGGTTACCATATATCCCAACTTCAGTCGGACACAGATCTACCTCAT 49882

Qy	133	CGGG--GGGGACCTGGGGCCCTTTTAAACCCTGCTTTACCCGTTGGAAAGTGCCTCTGGGCTG	190
Db	49881	TGGGGTGGGTACCTGGGGCCCTTTTAAACCCTGCTTTACCAAGTACCCCGTGGCTG	49822
Qy	191	GCATTAACCTGAAACAAGACAGAAATGTCGCCTGCTCCCTCCAGAGTGGATGGATGTA	250
Db	49821	GCAATTAACCTGAAACAAGACAGAAATGTCGACTGCTCCCTCCGAGTGGATGAACATA	49762
Qy	251	GAAGAAGTTGGAGAAGATGAGGATCATGAACGAAAGAGAAACATTTT	298
Db	49761	GAAGAAGTTGGAGAAGATGAGGATCATGAATGAAGGAAGAAACATTTT	49714
RESULT 10			
AX336730		454 bp	DNA
LOCUS	AX336730	Sequence 7239 from Patent WO0194629.	linear
DEFINITION	AX336730		
ACCESSION	AX336730.1	GI:18127449	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
		Homo sapiens (human)	
REFERENCE			
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G., Horrihan, S., Soppet, D.R. and Weaver, Z. Cancer gene determination and therapeutic screening using signature gene sets	
JOURNAL		Patent: WO 0194629-A 7239, 13-DEC-2001; Avalon Pharmaceuticals (US)	
FEATURES			
source		Location/Qualifiers	
		1..454	
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
BASE COUNT	128 a	113 c	116 g 97 t
ORIGIN			
Query Match		74.5%; Score 223.4; DB 6; Length 454;	
Best Local Similarity		99.5%; Pred. No. 3e-48;	
Matches	224; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
Qy	76	CGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGCTGGACAGAGATCTACCTCATCGG	135
Db	2	CGATAAGGAGCTGGTTACCATTTATCCCAACTTCAGCTGGACAGAGATCTACCTCATCGG	61
Qy	136	GGGGGACCTGGGGCCCTTTTAAACCCTGGTTTACCCGTTGGAAAGTGCCTCTGGCTGGCGAT	195
Db	62	GGGGGACCTGGGGCCCTTTTAAACCCTGGTTTACCCGTTGGAAAGTGCCTCTGGCTGGCGAT	121
Qy	196	TAACTTGAACAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGGATGGATGATAGAAA	255
Db	122	TAACTTGAACAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGGATGGATGATAGAAA	181
Qy	256	GTGGAGAGATGAGGATCATGAACGAAAGGAGAAACATTTTAC	300
Db	182	GTGGAGAGATGAGGATCATGAACGAAAGGAGAAACATTTTAC	226
RESULT 11			
AB097169		933 bp	DNA
LOCUS	AB097169	Xenopus laevis Psf2 gene, complete cds.	linear
DEFINITION	AB097169		
ACCESSION	AB097169.1	GI:29365480	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
		Xenopus laevis (African clawed frog)	
		Xenopus laevis	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.	

REFERENCE  
AUTHORS Kubota,Y., Takase,Y., Komori,Y., Hashimoto,Y., Arata,T., Kamimura,Y., Araki,H. and Takisawa,H.  
TITLE A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication  
JOURNAL Genes Dev. 17 (9), 1141-1152 (2003)  
MEDLINE 22615695  
PUBMED 12730133  
REFERENCE  
AUTHORS Kubota,Y. and Takisawa,H.  
TITLE Direct Submission  
JOURNAL Submitted (28-NOV-2002) Yumiko Kubota, Osaka University, Department of Biology, Graduate School of Science; Machikaneyamacho 1-1, Toyonaka, Osaka 560-0043, Japan  
(E-mail:ykubota@bio.sci.osaka-u.ac.jp, Tel:81-6-6850-5554, Fax:81-6-6850-5554)

FEATURES  
source Location/Qualifiers  
1. .933  
/organism="Xenopus laevis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:8355"  
21. 578  
/gene="Psf2"  
21. .578  
/function="essential for DNA replication"  
/codon\_start=1  
/product="Psf2"  
/protein\_id="BAC66459.1"  
/db\_xref="GI:29365481"  
/translation="MDASEVEFLAEKEQVTVIPNFSLDKVLIGGDLGPFNPSLPVEV PLWALNLRQKQKRVIPPEWMDVKLEAIRQERREFTFPMPSPIYMLKLLLNH AADNIPKADEIRTLVKDITWDTRIAKLRISADSVFKQEAHAKLDNLTLMEINTIGTFE TESLNHYMLRTSLQNPBEGSQDY"  
BASE COUNT 282 a 200 c 208 g 243 t  
ORIGIN

Query Match 55.8%; Score 167.4; DB 5; Length 933;  
Best Local Similarity 78.2%; Pred. No. 1.8e-33;  
Matches 201; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 44 GCATGAGCGCTGCCAGAGTCGAATTCCTCGCGGAGAGGAGCGTGGTTACCATATACCCC 103  
Db 18 GTACATGATGCGCTCTGAGGTGAGTTCCTGGGTGAGAGGAGCAAGTACCGCTCATCCCC 77  
QY 104 AACTTCAGTCTGCACAAATATCACTCATCGGGGGGAGCTGGGGCTTTTAAACCCCTGTT 163  
Db 78 AACTTAGTCTGGACAAGATCTATCTCATTTGGGGGTGACTTGGGCCCTTTTAACTCCTAGT 137  
QY 164 TTACCCGTGGAAGTGCCTGCTGGCTGGCGGATTAACCTGAAACAAAGACAGAAATGTCG 223  
Db 138 TTACCTGTTGAGGTTCTCTGTTGGTGGCCATCACTGAACAAAGGAGCAAAATGTCGA 197  
QY 224 CTGCTCCCTCCAGATGGATGATGATAGAAAAGTTGGAGAAGATGAGGATCATGAACGA 283  
Db 198 ATTTGTCCTCCAGATGGATGATGATGAGAAAAGTTGAGAACTAGAGGCTATCCGAGACCAAGAACGC 257  
QY 284 AAGGAGAAACTTTTAC 300  
Db 258 AGAGGAGACGCTTTAC 274

RESULT 12  
AC009068/c  
LOCUS AC009068 160356 bp DNA linear PRI 27-APR-2001  
DEFINITION Homo sapiens chromosome 16 clone RP11-314K3, complete sequence.  
ACCESSION AC009068  
VERSION AC009068.10 GI:13811892  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 160356)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (27-APR-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Apr 27, 2001 this sequence version replaced gi:12000278.  
COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.6% of Sequence;  
Estimated Total Number of Errors is 0.3.  
STS Content:  
WI-11796 G24048  
SHGC-37143 G30481.

FEATURES  
source Location/Qualifiers  
1. .160356  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-314K3"  
BASE COUNT 37469 a 43134 c 43308 g 36445 t  
ORIGIN

Query Match 41.7%; Score 125.2; DB 9; Length 160356;  
Best Local Similarity 94.2%; Pred. No. 3.7e-22;  
Matches 130; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 13 GTCGGTGAGGCTCTGCGCTGCAGCTCGCGCGGCATGGAGCTGCCGAGTCCGAATTCCT 72  
Db 60646 GACCGTGAGGCTCTGCGCTGCAGCTCGCGCGGCATGGAGCTGCCGAGTCCGAATTCCT 60587  
QY 73 GCGCGAGAAGAGCTGGTTACCATATATCCCACTTTCAGTCTGCACAAAGATCTACCTCAT 132  
Db 60586 GCGCGAGAAGAGCTGGTTACCATATATCCCACTTTCAGTCTGCACAAAGATCTACCTCAT 60527  
QY 133 CGGGGGGAGCTGGGGGCC 150  
Db 60526 CGGGGTGAGGCTGGGGCC 60509

RESULT 13  
AC123908/c  
LOCUS AC123908 185321 bp DNA linear PRI 28-MAR-2003  
DEFINITION Homo sapiens chromosome 16 clone CTD-2542L18, complete sequence.  
ACCESSION AC123908  
VERSION AC123908.3 GI:29336199  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (05-JUN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 185321)

AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (06-SEP-2002) Production Sequencing Facility, DOE Joint  
REFERENCE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
AUTHORS 4 (bases 1 to 185321)  
DOE Joint Genome Institute, Stanford Human Genome Center and Los  
Alamos National Laboratory.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Mar 28, 2003 this sequence version replaced gi:22748376.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center and Los Alamos  
National Laboratory  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.  
FEATURES  
source 1..185321  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2542L18"  
BASE COUNT 40674 a 50904 c 51869 g 41874 t  
ORIGIN  
Query Match 41.7%; Score 125.2; DB 9; Length 185321;  
Best Local Similarity 94.2%; Pred. No. 3.7e-22;  
Matches 130; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 13 GTCGGTAGGCTCTGGCTGAGCTGCGCGCCATGCGAGCTGCGAGGTCGAATTCCT 72  
Db 105811 GACCGTAGGCTCTGGCTGAGCTGCGCGCCATGCGAGCTGCGAGGTCGAATTCCT 105752  
QY 73 CGCCGAGAGGAGCTGTTACCATATCCCACTTCCAGTGTGACAGATCTACCTCAT 132  
Db 105751 CGCCGAGAGGAGCTGTTACCATATCCCACTTCCAGTGTGACAGATCTACCTCAT 105692  
QY 133 CGGGGGGAGCCTGGGGCC 150  
Db 105691 CGGGGTGAGGCTGGGCC 105674  
RESULT 14  
AC137771/c 245210 bp DNA linear HTG 03-DEC-2002  
LOCUS AC137771  
DEFINITION Homo sapiens chromosome 16 clone CTD-2139A24, WORKING DRAFT  
SEQUENCE, 45 unordered pieces.  
ACCESSION AC137771  
VERSION AC137771.1 GI:26006513  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 245210)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 245210)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-DEC-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 680949

Center clone name: CITB-HI\_2139A24  
-----  
Summary Statistics  
Consensus quality: 221831 bases at least Q40  
Consensus quality: 226530 bases at least Q30  
Consensus quality: 229658 bases at least Q20  
Estimated insert size: 170000; agarose-fp estimation  
Quality coverage: 12.32 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 12.32 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1084: contig of 1084 bp in length  
\* 1085 1184: gap of unknown length  
\* 1185 2248: contig of 1064 bp in length  
\* 2249 2348: gap of unknown length  
\* 2349 3435: contig of 1087 bp in length  
\* 3436 3535: gap of unknown length  
\* 3536 4637: contig of 1102 bp in length  
\* 4638 4737: gap of unknown length  
\* 4738 5816: contig of 1079 bp in length  
\* 5817 5916: gap of unknown length  
\* 5917 6951: contig of 1035 bp in length  
\* 6952 7051: gap of unknown length  
\* 7052 8599: contig of 1548 bp in length  
\* 8600 8699: gap of unknown length  
\* 8700 10508: contig of 1809 bp in length  
\* 10509 10608: gap of unknown length  
\* 10609 11738: contig of 1130 bp in length  
\* 11739 11838: gap of unknown length  
\* 11839 13754: contig of 1916 bp in length  
\* 13755 13854: gap of unknown length  
\* 13855 14996: contig of 1142 bp in length  
\* 14997 15096: gap of unknown length  
\* 15097 16852: contig of 1756 bp in length  
\* 16853 16952: gap of unknown length  
\* 16953 18502: contig of 1550 bp in length  
\* 18503 18602: gap of unknown length  
\* 18603 20736: contig of 2134 bp in length  
\* 20737 20836: gap of unknown length  
\* 20837 22243: contig of 1407 bp in length  
\* 22244 22343: gap of unknown length  
\* 22344 24310: contig of 1967 bp in length  
\* 24311 24410: gap of unknown length  
\* 24411 26675: contig of 2265 bp in length  
\* 26676 26775: gap of unknown length  
\* 26776 28257: contig of 1482 bp in length  
\* 28258 28357: gap of unknown length  
\* 28358 29668: contig of 1311 bp in length  
\* 29669 29768: gap of unknown length  
\* 29769 31503: contig of 1735 bp in length  
\* 31504 31603: gap of unknown length  
\* 31604 33962: contig of 2359 bp in length  
\* 33963 34062: gap of unknown length  
\* 34063 36099: contig of 2037 bp in length  
\* 36100 38832: contig of 2633 bp in length  
\* 38833 38932: gap of unknown length  
\* 38933 41464: contig of 2532 bp in length  
\* 41465 41564: gap of unknown length  
\* 41565 44085: contig of 2521 bp in length  
\* 44086 44185: gap of unknown length  
\* 44186 47028: contig of 2843 bp in length  
\* 47029 47128: gap of unknown length  
\* 47129 49395: contig of 2267 bp in length  
\* 49396 49495: gap of unknown length  
\* 49496 51435: contig of 1940 bp in length  
\* 51435



may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GMCE

Center clone name: CH230-247G13

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 198554 bases at least Q40

Consensus quality: 200661 bases at least Q30

Consensus quality: 201900 bases at least Q20

Estimated insert size: 210626; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 1 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 207307: contig of 207307 bp in length.

#### FEATURES

##### source

1. 207307

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-247G13"

87039..177061

/note="clone\_boundary"

clone\_end:Sp6

site:Mbol

end\_sequence:Bzl68551"

150202..150357

/note="clone\_boundary"

clone\_end:T7

site:Mbol

end\_sequence:Bzl68550"

BASE COUNT 43941 a 54080 c 55681 g 48886 t 4719 others

##### ORIGIN

Query Match 32.3%; Score 97; DB 2; Length 207307;

Best Local Similarity 87.6%; Pred. No. 1e-14; Mismatches 15; Indels 0; Gaps 0;

Matches 106; Conservative 0;

QY 136 GGGGACCTGGGGCCCTTTTAAACCTGGTTTACCCGTGGTGAAGTCCCTGTGGCTGGCGAT 195

|||||

Db 205359 GGGGACCTGGGGCCCTTCAACCCCGCTTACCGTGGACGTGCCCTGTGGCTGGCCAT 205300

|||||

QY 196 TAACCTGAACAACAGACAGAAATGCGCTGTCTCCCTCCAGAGTGGATGGATGAGAAAA 255

|||||

Db 205299 TAACCTGAACAACAGACAGAAATGCGCTGTCTCCCTCCAGAGTGGATGGATGAGAAAA 255

|||||

QY 256 G 256

Db 205239 G 205239

Search completed: October 14, 2003, 17:02:29  
Job time : 1481 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 13:05:16 ; Search time 36.6667 Seconds  
(without alignments)  
3611.317 Million cell updates/sec

Title: US-09-854-124-5  
Perfect score: 300  
Sequence: 1 acgaatccgagccctggtc.....tgaggtactcgtggttctgg 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2.6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2.6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	11.7	2241	2	US-08-838-219B-20
2	35	11.7	2241	3	US-09-233-336A-20
3	35	11.7	2241	3	US-09-233-752A-20
4	35	11.7	2241	3	US-09-402-036-20
5	35	11.7	2241	4	US-09-904-226-20
6	35	11.7	2370	2	US-08-838-219B-19
7	35	11.7	2370	3	US-09-233-336A-19
8	35	11.7	2370	3	US-09-233-752A-19
9	35	11.7	2370	3	US-09-402-036-19
10	35	11.7	2370	4	US-09-904-226-19
11	35	11.7	2403	1	US-08-471-033-30
12	35	11.7	2403	2	US-08-471-044-30
13	35	11.7	2403	2	US-08-463-483A-30
14	35	11.7	2403	2	US-08-471-046A-30
15	35	11.7	2403	2	US-08-470-566B-30
16	35	11.7	2403	2	US-08-838-219B-7
17	35	11.7	2403	2	US-08-469-334-30
18	35	11.7	2403	3	US-09-300-529-30
19	35	11.7	2403	3	US-09-233-336A-7
20	35	11.7	2403	3	US-09-233-752A-7
21	35	11.7	2403	3	US-09-402-036-7
22	35	11.7	2403	4	US-09-904-226-7
23	31.4	10.5	1296	4	US-09-252-991A-7149
24	31.4	10.5	1419	4	US-09-252-991A-7347
25	31.4	10.5	2118	4	US-09-016-434-1304
26	31.4	10.5	2748	4	US-09-252-991A-7208
27	31.4	10.5	2831	2	US-08-906-713-1

Sequence 43, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 33, Appli  
Sequence 3, Appli  
Sequence 5, Appli  
Sequence 31, Appli  
Sequence 21, Appli  
Sequence 2, Appli  
Sequence 1192, Ap  
Patent No. 5472691  
Sequence 1, Appli  
Patent No. 5472691  
Sequence 12012, A  
Sequence 11975, A  
Sequence 11906, A  
Patent No. 5472691  
Sequence 20, Appli

ALIGNMENTS

RESULT 1  
US-08-838-219B-20  
; Sequence 20, Application US/08838219B  
; Patent No. 5877012  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: A No. 5877012el Class of Proteins for the  
; TITLE OF INVENTION: Control of Plant Pests  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/838,219B  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/463,483  
; FILING DATE: 06-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pace, Gary M.  
; REGISTRATION NUMBER: 40,403  
; REFERENCE/DOCKET NUMBER: CGC 1925  
; TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2241 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA encoding
; DESCRIPTION: VIP3A(c)"
; HYPOTHETICAL: NO
; US-08-838-219B-20

Query Match      11.7%; Score 35; DB 2; Length 2241;
Best Local Similarity 59.6%; Pred. No. 0.14;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 77 GAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACCAG 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1950 GAGCCAGAAACGGCAGCAGGCGCTGGGGCGACAACCTTCATCTCCTCTGGAGATCAGCCCGAG 2009

QY 137 CGGACTTTCCTCACACAAGCGCTCAACCATGTGTACAA 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2010 CGAGAGCTGTGACCCGGAGCTGATCAACACCAACAA 2048

RESULT 2
US-09-233-336A-20
; Sequence 20, Application US/09233336A
; Patent No. 6107279
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nallini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: A No. 6107279el Class of Proteins for the
; TITLE OF INVENTION: Control of Plant Pests
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/233,336A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/838,219
; FILING DATE:
; PRIORITY DATE:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:

;
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1925
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2241 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA encoding
; DESCRIPTION: VIP3A(c)"
; HYPOTHETICAL: NO
; US-09-233-336A-20

Query Match      11.7%; Score 35; DB 3; Length 2241;
Best Local Similarity 59.6%; Pred. No. 0.14;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 77 GAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACCAG 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1950 GAGCCAGAAACGGCAGCAGGCGCTGGGGCGACAACCTTCATCTCCTCTGGAGATCAGCCCGAG 2009

QY 137 CGGACTTTCCTCACACAAGCGCTCAACCATGTGTACAA 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2010 CGAGAGCTGTGACCCGGAGCTGATCAACACCAACAA 2048

RESULT 3
US-09-233-752A-20
; Sequence 20, Application US/09233752A
; Patent No. 6137033
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nallini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: A No. 6137033el Class of Proteins for the
; TITLE OF INVENTION: Control of Plant Pests
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/233,752A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/838,219
; FILING DATE:
; PRIORITY DATE:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
```



RESULT 6  
US-08-838-219B-19

RESULT 6  
US-08-838-219B-19

; Sequence 19, Application US/08838219B  
; Patent No. 5877012

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J

TITLE OF INVENTION: A No. 5877012el Class of Proteins for the  
TITLE OF INVENTION: Control of Plant Pests

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/838,219B

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/463,483

FILING DATE: 06-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Pace, Gary M.

REGISTRATION NUMBER: 40,403

REFERENCE/DOCKET NUMBER: CGC 1925

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8582

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 2370 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "synthetic DNA encoding

DESCRIPTION: VIP3A(b)"

HYPOTHETICAL: NO

US-08-838-219B-19

Query Match 11.7%; Score 35; DB 2; Length 2370;

Best Local Similarity 59.6%; Pred. No. 0.15;

Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 77 GAGACAGCAGGACATGCCAGCTGGATACTTGCACCTTGATGGAGATCAACACCAG 136

Db 1950 GAGCAGACGGCAGCAGGCGCTGGGCGGACAACTTCATCATCTCGGAGATCAGCCGAG 2009

QY 137 CGGACATTTCTCTACACAAGCGCTCAACCCATCATGACAA 175

Db 2010 CGAGAGCTGCTGAGCCCGGAGCTGTGATCAACACCAACAA 2048

RESULT 7

US-09-233-336A-19

; Sequence 19, Application US/09233336A

; Patent No. 6107279

; GENERAL INFORMATION:

APPLICANT: Warren, Gregory W

APPLICANT: Koziel, Michael G

APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J

APPLICANT: Carr, Brian

APPLICANT: Desai, Nalini M

APPLICANT: Kostichka, N. Kristy

APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J

TITLE OF INVENTION: A No. 6107279el Class of Proteins for the

TITLE OF INVENTION: Control of Plant Pests

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/233,336A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/838,219

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Pace, Gary M.

REGISTRATION NUMBER: 40,403

REFERENCE/DOCKET NUMBER: CGC 1925

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8582

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 2370 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "synthetic DNA encoding

DESCRIPTION: VIP3A(b)"

HYPOTHETICAL: NO

US-09-233-336A-19

Query Match 11.7%; Score 35; DB 3; Length 2370;

Best Local Similarity 59.6%; Pred. No. 0.15;

Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 77 GAGACAGCAGGACATGCCAGCTGGATACTTGCACCTTGATGGAGATCAACACCAG 136

Db 1950 GAGCAGACGGCAGCAGGCGCTGGGCGGACAACTTCATCATCTCGGAGATCAGCCGAG 2009



```
; APPLICANT: Koziel, Michael
; APPLICANT: Nye, Gordon
; TITLE OF INVENTION: Plant Pest Control
; FILE REFERENCE: S-21284D
; CURRENT APPLICATION NUMBER: US/09/904,226
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/EP98/01952
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 08/838,219
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,263
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,265
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/463,483
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/314,594
; PRIOR FILING DATE: 1994-09-09
; PRIOR APPLICATION NUMBER: 08/218,018
; PRIOR FILING DATE: 1994-03-24
; PRIOR APPLICATION NUMBER: 08/037,057
; PRIOR FILING DATE: 1993-03-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; OTHER INFORMATION: encoding VIP3A(b)
US-09-904-226-19

Query Match      11.7%; Score 35; DB 4; Length 2370;
Best Local Similarity 59.6%; Pred. No. 0.15;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 77 GAGACAGCAGGAGCGACATGCCAAGCTGGGATTAACCTTGACCTTGATGGAGATCAACACAG 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1950 GAGCCAGAACGGCGACGAGCGCTGGGGCGACAACTTCATCATCTCTGGAGATCAGCCCGAG 2009
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 CGGACTTTCTCTACACAAAGCGCTCAACACCATGTACAA 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2010 CGAGAGCTGCTGAGCCCGGAGCTGATCAACACCAACAA 2048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-471-033-30
; Sequence 30, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11..2389
; OTHER INFORMATION: /note= "maize optimized DNA"
; OTHER INFORMATION: sequence encoding VIP3A(a)"
US-08-471-033-30

Query Match      11.7%; Score 35; DB 1; Length 2403;
Best Local Similarity 59.6%; Pred. No. 0.15;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 77 GAGACAGCAGGAGCGACATGCCAAGCTGGGATTAACCTTGACCTTGATGGAGATCAACACAG 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1966 GAGCCAGAACGGCGACGAGCGCTGGGGCGACAACTTCATCATCTCTGGAGATCAGCCCGAG 2025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 CGGACTTTCTCTACACAAAGCGCTCAACACCATGTACAA 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2026 CGAGAGCTGCTGAGCCCGGAGCTGATCAACACCAACAA 2064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-08-471-044-30
; Sequence 30, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2403 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11..2389  
OTHER INFORMATION: /note= "maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIP3A(a)"  
US-08-471-044-30

Query Match 11.7%; Score 35; DB 2; Length 2403;  
Best Local Similarity 59.6%; Pred. No. 0.15;  
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 77 GAGACAGCAGGAGCAGCATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACCAG 136  
Db 1966 GAGCCAGAACGGCGACGAGGCGCTGGGGCGACAACTTCATCATCTCTGGAGATCAGCCCGAG 2025  
Qy 137 CGGAGCTTTCTCTACACAAGCGCTCAACACCATGTACAA 175  
Db 2026 CGAGAAGCTGCTGAGCCCGGAGCTGATCAACACCAACAA 2064

RESULT 13  
US-08-463-483A-30  
Sequence 30, Application US/08463483A  
Patent No. 5849870  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,483A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Spruill, W. Murray  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8615  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2403 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11..2389  
OTHER INFORMATION: /note= "maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIP3A(a)"  
US-08-463-483A-30

Query Match 11.7%; Score 35; DB 2; Length 2403;  
Best Local Similarity 59.6%; Pred. No. 0.15;  
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 77 GAGACAGCAGGAGCAGCATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACCAG 136  
Db 1966 GAGCCAGAACGGCGACGAGGCGCTGGGGCGACAACTTCATCATCTCTGGAGATCAGCCCGAG 2025  
Qy 137 CGGAGCTTTCTCTACACAAGCGCTCAACACCATGTACAA 175  
Db 2026 CGAGAAGCTGCTGAGCCCGGAGCTGATCAACACCAACAA 2064

RESULT 14  
US-08-471-046A-30  
Sequence 30, Application US/08471046A  
Patent No. 5866326  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy

```

; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
; TITLE OF INVENTION: Protein Genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5860326artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,046A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11..2389
; OTHER INFORMATION: /note= "maize optimized DNA"
; OTHER INFORMATION: sequence encoding VIP3A(a)"
; US-08-471-046A-30
;
; Query Match 11.7%; Score 35; DB 2; Length 2403;
; Best Local Similarity 59.6%; Pred. No. 0.15;
; Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
;
; QY 77 GAGACAGCAGGAGGACATGCCAAGCTGGATACCTTGATGGATGAGATCAACACCAG 136
; Db 1966 GAGCCAGACGGCGAGCGAGCGCTGGGGCGACAACTTCATCTCCTGGATCAGCCCGAG 2025
;
; QY 137 CGGGACTTTCCTCACACAAGCGCTCAACCCATGTACAA 175
; Db 2026 CGAAGAGCTGCTAGCGCCGAGCTGATCAACACCAACNA 2064
;
; RESULT 15
; US-08-470-566B-30
; Sequence 30, Application US/08470566B
; Patent No., 5872212
```

```

; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5872212artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,566B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11..2389
; OTHER INFORMATION: /note= "maize optimized DNA"
; OTHER INFORMATION: sequence encoding VIP3A(a)"
; US-08-470-566B-30
;
; Query Match 11.7%; Score 35; DB 2; Length 2403;
; Best Local Similarity 59.6%; Pred. No. 0.15;
; Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
;
; QY 77 GAGACAGCAGGAGGACATGCCAAGCTGGATACCTTGATGGATGAGATCAACACCAG 136
; Db 1966 GAGCCAGACGGCGAGCGAGCGCTGGGGCGACAACTTCATCTCCTGGATCAGCCCGAG 2025
;
; QY 137 CGGGACTTTCCTCACACAAGCGCTCAACCCATGTACAA 175
; Db 1966 GAGCCAGACGGCGAGCGAGCGCTGGGGCGACAACTTCATCTCCTGGATCAGCCCGAG 2025
```

Db 2026 CGAGAGCTGTGAGCCCGGAGCTGATCAACACCAACAA 2064

Search completed: October 14, 2003, 17:04:21  
Job time : 38.6667 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 14:17:11 ; Search time 143.333 Seconds  
(without alignments)  
5431.000 Million cell updates/sec

Title: US-09-854-124-5

Perfect score: 300

Sequence: 1 acgaatacggaccctggtc.....tgaggtaactcgtgtctgg 300

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications, NA.\*  
1: /cgn2.6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2.6/ptodata/2/pubpna/FCI\_NEW\_PUB.seq.\*  
3: /cgn2.6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2.6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2.6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2.6/ptodata/2/pubpna/PTUS\_PUBCOMB.seq.\*  
7: /cgn2.6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2.6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2.6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2.6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2.6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2.6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2.6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2.6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2.6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2.6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2.6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	300	100.0	300	9	US-09-854-124-5 Sequence 5, Appli
2	300	100.0	1210	9	US-09-925-701-214 Sequence 214, App
3	239	79.7	484	11	US-09-918-995-19938 Sequence 19938, A
4	214	71.5	200	9	US-09-854-124-7 Sequence 7, Appli
5	199	66.3	382	10	US-09-796-692-9460 Sequence 9460, Ap
6	199	66.3	382	14	US-10-040-862-9460 Sequence 9460, Ap
7	194	64.7	597	10	US-09-796-692-9407 Sequence 9407, Ap
8	194	64.7	597	14	US-10-040-862-9407 Sequence 9407, Ap
9	66	22.0	362	10	US-09-920-300A-682 Sequence 682, App
10	66	22.0	362	10	US-09-998-538-2284 Sequence 2284, Ap
11	66	22.0	362	12	US-10-099-946-682 Sequence 682, App
12	66	22.0	362	13	US-10-033-528-682 Sequence 682, App
13	39.2	13.1	820	13	US-10-027-632-151845 Sequence 151845, A
14	39.2	13.1	820	13	US-10-027-632-151846 Sequence 151846, A
15	32.8	10.9	468	9	US-09-854-761-789 Sequence 789, App
16	32.8	10.9	737	9	US-09-854-761-17572 Sequence 17572, A

17	32.4	10.8	1188	14	US-10-156-761-5491 Sequence 5491, Ap
c 18	32.4	10.8	9025608	14	US-10-156-761-1 Sequence 1, Appli
c 19	31.6	10.5	650	13	US-10-027-632-246523 Sequence 246523, A
20	31.6	10.5	831	14	US-10-156-761-4064 Sequence 4064, Ap
21	31.4	10.5	360	10	US-09-920-455-112 Sequence 112, App
22	31.4	10.5	517	10	US-09-998-598-1387 Sequence 1387, Ap
23	31.4	10.5	596	13	US-10-027-632-50355 Sequence 50355, A
24	31.4	10.5	596	13	US-10-027-632-50356 Sequence 50356, A
c 25	31.4	10.5	1722	10	US-09-912-672A-3 Sequence 3, Appli
c 26	31.4	10.5	2478	12	US-10-063-735-163 Sequence 163, App
c 27	31.4	10.5	2478	12	US-10-063-526-163 Sequence 163, App
c 28	31.4	10.5	2478	12	US-10-063-586-163 Sequence 163, App
c 29	31.4	10.5	2478	12	US-10-063-510-163 Sequence 163, App
c 30	31.4	10.5	2478	12	US-10-063-514-163 Sequence 163, App
c 31	31.4	10.5	2478	12	US-10-063-516-163 Sequence 163, App
c 32	31.4	10.5	2478	12	US-10-063-523-163 Sequence 163, App
c 33	31.4	10.5	2478	12	US-10-063-527-163 Sequence 163, App
c 34	31.4	10.5	2478	12	US-10-063-528-163 Sequence 163, App
c 35	31.4	10.5	2478	12	US-10-063-529-163 Sequence 163, App
c 36	31.4	10.5	2478	12	US-10-063-536-163 Sequence 163, App
c 37	31.4	10.5	2478	12	US-10-063-540-163 Sequence 163, App
c 38	31.4	10.5	2478	12	US-10-063-546-163 Sequence 163, App
c 39	31.4	10.5	2478	12	US-10-063-562-163 Sequence 163, App
c 40	31.4	10.5	2478	12	US-10-063-564-163 Sequence 163, App
c 41	31.4	10.5	2478	12	US-10-063-565-163 Sequence 163, App
c 42	31.4	10.5	2478	12	US-10-063-568-163 Sequence 163, App
c 43	31.4	10.5	2478	12	US-10-063-570-163 Sequence 163, App
c 44	31.4	10.5	2478	12	US-10-063-577-163 Sequence 163, App
c 45	31.4	10.5	2478	12	US-10-063-577-163 Sequence 163, App

#### ALIGNMENTS

#### RESULT 1

US-09-854-124-5  
; Sequence 5, Application US/09854124  
; Patent No. US20020076735A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Escobedo, Jaime  
; APPLICANT: Innis, Michael A.  
; APPLICANT: Garcia, Pablo Dominiguez  
; APPLICANT: Sudduth-Klinger, Julie  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Giese, Klaus  
; APPLICANT: Randazzo, Filippo  
; APPLICANT: Kennedy, Giulia C.  
; APPLICANT: Pot, David  
; APPLICANT: Kassam, Altaf  
; APPLICANT: Lamsom, George  
; APPLICANT: Drmanac, Radoje  
; APPLICANT: Crkvenjakov, Radomir  
; APPLICANT: Dickson, Mark  
; APPLICANT: Drmanac, Snezana  
; APPLICANT: Labat, Ivan  
; APPLICANT: Leshkowitz, Dena  
; APPLICANT: Kita, David  
; APPLICANT: Garcia, Veronica  
; APPLICANT: Jones, Lee William  
; APPLICANT: Stache-Crain, Birgit  
; TITLE OF INVENTION: Diagnostic and Therapeutic Methods Using Molecules Differentially Expressed in Cancer Cells  
; FILE REFERENCE: 2300-1490  
; CURRENT APPLICATION NUMBER: US/09/854,124  
; CURRENT FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: 09/400,947  
; PRIOR FILING DATE: 1999-09-22  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 300  
; TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-854-124-5

Query Match
Best Local Similarity 100.0%; Score 300; DB 9; Length 300;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGACCTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 60
Db 1 ACGAAATCCGACCTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 60
QY 61 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 120
Db 61 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 120
QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACACAAGCGCTCAACCAATGTACAACTCC 180
Db 121 TGGAGATCAACACAGCGGGACTTTCCTCACACAAGCGCTCAACCAATGTACAACTCC 180
QY 181 GCAGAACTCCAGCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCGCTGGT 240
Db 181 GCAGAACTCCAGCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCGCTGGT 240
QY 241 GCAGCGGCTTGTGCGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGCTTGG 300
Db 241 GCAGCGGCTTGTGCGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGCTTGG 300

RESULT 2
US-09-925-301-214
; Sequence 214, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; FILE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCI/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 214
; LENGTH: 1210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-214

Query Match
Best Local Similarity 100.0%; Score 300; DB 9; Length 1210;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGACCTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 60
Db 412 ACGAAATCCGACCTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 471
QY 61 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 120
Db 472 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 531
QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACACAAGCGCTCAACCAATGTACAACTCC 180
Db 532 TGGAGATCAACACAGCGGGACTTTCCTCACACAAGCGCTCAACCAATGTACAACTCC 591
QY 181 GCAGAACTCCAGCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCGCTGGT 240
Db 592 GCAGAACTCCAGCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCGCTGGT 651
QY 241 GCAGCGGCTTGTGCGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGCTTGG 300
Db 652 GCAGCGGCTTGTGCGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGCTTGG 711
```

```
RESULT 3
US-09-918-995-19938
; Sequence 19938, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19938
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-19938
```

```
Query Match
Best Local Similarity 100.0%; Score 239; DB 11; Length 484;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGACCTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 60
Db 245 ACGAAATCCGACCTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 304
QY 61 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 120
Db 305 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCGTATAGCCAAACTCCGAGTGT 364
QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACACAAGCGCTCAACCAATGTACAACTCC 180
Db 365 TGGAGATCAACACAGCGGGACTTTCCTCACACAAGCGCTCAACCAATGTACAACTCC 424
QY 181 GCAGAACTCCAGCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCGCTGG 239
Db 425 GCAGAACTCCAGCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCGCTGG 483
```

```
RESULT 4
US-09-854-124-7
; Sequence 7, Application US/09854124
; Patent No. US20020076735A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: Diagnostic and Therapeutic Methods Using
; MOLECULES DIFFERENTIALLY EXPRESSED IN CANCER CELLS
```

```

; FILE REFERENCE: 2300-1490
; CURRENT APPLICATION NUMBER: US/09/854,124
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/400,947
; PRIOR FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-854-124-7

Query Match      71.5%; Score 214.4; DB 9; Length 300;
Best Local Similarity 94.6%; Pred. No. 6.4e-62;
Matches 244; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ACGAATCCGACCGCTGTTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGCACCTTGA 120
   |||||
Db 31 ACGAATCCGACCGCTGTTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGCACCTTGA 90
   |||||

QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGCACCTTGA 120
   |||||
Db 91 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGCACCTTGA 150
   |||||

QY 121 TGGAGATCAACAGCAGCGGACTTTCCTCACACAGCGCTCAACCCACATGTACAACTCC 180
   |||||
Db 151 TGGAGATCAACAGCAGCGGACTTTCCTCACACAGCGCTCAACCCACATGTACAACTCC 210
   |||||

QY 181 GCAGGAACCTCCAGCTCTGGAGACTCTGAGACTCAGCTCAGGACTCTAGAGAAAGCGCTGT 240
   |||||
Db 211 GCAGGAACCTCCAGCTCTGGAGACTCTGAGAGACCTCAG--CTAGGACTCTA-AAAAAGGCGCTGT 267
   |||||

QY 241 GCAGGCGGCTTGTCTGGGG 258
   |||||
Db 268 GCAGGCGCTTGTGGGG 285

RESULT 5
US-09-796-692-9460/c
; Sequence 9460, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077,001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04

; FILE REFERENCE: 2300-1490
; CURRENT APPLICATION NUMBER: US/09/854,124
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/400,947
; PRIOR FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9460
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (72)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (80)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (89)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (95)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (122)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (129)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (131)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (155)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (173)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (222)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (235)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (236)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (237)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (245)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (251)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (261)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (305)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (312)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (321)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (334)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
```

```
; LOCATION: (352)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (354)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (373)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (375)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-9460

Query Match      66.3%; Score 199; DB 10; Length 382;
Best Local Similarity 94.4%; Pred. No. 1e-56;
Matches 221; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY      3  GAAATCCGGACCCCTGGTCAAGCATATGCGACACTCTGTATAGCCAACTCCGAGTGTCT 62
DB      234  GAAATCCGGACCCCTGGTCAAGCATATGCGACACTCTGTATAGCCAACTCCGAGTGTCT 175

QY      63  GCTGACAGCTTTGTGAGACA-GCAGGAGGCACATGCCAAGCTGGATAACTTGGACCTTGAT 121
DB      174  GNTGACAGCTTTGTGAGACNCGCAGGAGGCACATGCCAAGCTGNANAACTTGNCCCTTGAT 115

QY      122  GGAGATCAACACAGCGGGACTTTCCTCACACA-AGCGCTCAACACCATGTACAAACTCC 180
DB      114  GGAGATCAACACAGCGGGGNCCTTTCNTCACACATNGCGCTCANGCAGCATGTACAAACTCC 55

QY      181  GCAGCAACTCCAGCCTCTGGAGAGTACTCACTCTCAGGACTCTAGAGAAAGG 234
DB      54  GCAGCAACTCCAGCCTCTGGAGAGTACTCACTCTCAGGACTTNTAGAGAAAGG 1

RESULT 6
US-10-040-862-9460/c
; Sequence 9460, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040.862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
```

```
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9460
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (72)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (80)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (89)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (95)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (122)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (129)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (131)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (155)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (173)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (222)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (235)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (236)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (237)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (245)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (251)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
```

```

; NAME/KEY: unsure
; LOCATION: (261)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (305)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (312)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (321)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (334)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (352)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (354)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (373)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (375)
; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-9460

```

Query Match	56.3%;	Score 199;	DB 14;	Length 382;
Best Local Similarity	94.4%;	Pred. No. le=56;		
Matches 221;	Conservative	0;	Mismatches 11;	Indels 2;
Gaps	2;			
QY	3	GAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGTCT	62	
Db	234	GAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGTCT	175	
QY	63	GCTGACAGCTTTGTGAGACA-GCAGGAGGCACATGCCAAGCTGGATAACTTGCACCTTGAT	121	
Db	174	GNTGACAGCTTTGTGAGACNCGCAGGAGGCACATGCCAAGCTGNANAACTTGNCTTGTAT	115	
QY	122	GGATGCAACACACAGCGGGACTTTCCTTCACACA-AGCGCTCAACCCACATGTACAAATCC	180	
Db	114	GGATGCAACACACAGCGGGNCTTTCNTCACACATNGCGCTCANGCCATGTACAAATCC	55	
QY	181	GCACGAACCTCCAGCCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGG	234	
Db	54	GCACGAACCTCCAGCCCTCTGGAGAGTACTCAGTCTCAGGACTTNTAGAGAAAGG	1	

RESULT 7  
US-09-796-692-4407  
; Sequence 4407, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannon, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077\_001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01

```

1 PRIOR APPLICATION NUMBER: 60/190,479
2 PRIOR FILING DATE: 2000-03-17
3 PRIOR APPLICATION NUMBER: 60/200,545
4 PRIOR FILING DATE: 2000-04-27
5 PRIOR APPLICATION NUMBER: 60/200,303
6 PRIOR FILING DATE: 2000-04-28
7 PRIOR APPLICATION NUMBER: 60/200,779
8 PRIOR FILING DATE: 2000-04-28
9 PRIOR APPLICATION NUMBER: 60/200,999
10 PRIOR FILING DATE: 2000-05-01
11 PRIOR APPLICATION NUMBER: 60/202,084
12 PRIOR FILING DATE: 2000-05-04
13 PRIOR APPLICATION NUMBER: 60/206,201
14 PRIOR FILING DATE: 2000-05-22
15 PRIOR APPLICATION NUMBER: 60/218,950
16 PRIOR FILING DATE: 2000-07-14
17 PRIOR APPLICATION NUMBER: 60/222,903
18 PRIOR FILING DATE: 2000-08-03
19 PRIOR APPLICATION NUMBER: 60/223,416
20 PRIOR FILING DATE: 2000-08-04
21 PRIOR APPLICATION NUMBER: 60/223,378
22 PRIOR FILING DATE: 2000-08-07
23 NUMBER OF SEQ ID NOS: 9597
24 SOFTWARE: FASTSEQ for Windows Version 3.0
25 SEQ ID NO 4407
26 LENGTH: 597
27 TYPE: DNA
28 ORGANISM: Homo sapiens
29 FEATURE:
30 NAME/KEY: unsure
31 LOCATION: (583)
32 OTHER INFORMATION: n=A,T,C or G
33 NAME/KEY: unsure
34 LOCATION: (592)
35 OTHER INFORMATION: n=A,T,C or G
36 US-09-796-692-4407

```

Query Match	64.7%	Score 194;	DB 10;	Length 597;
Best Local Similarity	99.0%	Pred. No. 5.7e-55;		
Matches 194;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

  

Qy	1	ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATATAGCCAAACTCCGAGTGT	60
Db	402	ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATATAGCCAAACTCCGAGTGT	461
Qy	61	CTGCTGACAGGCTTTGTGAGACAGCAGGAGGACACATGCCAAGCTGGATAAAGCTTGACCTTGA	120
Db	462	CTGCTGACAGGCTTTGTGAGACAGCAGGAGGACACATGCCAAGCTGGATAAAGCTTGACCTTGA	521
Qy	121	TGGAGATCAACACCAGCGGGGACITTCCTCACACAAGCGCTCAACCATATGTACAAATCC	180
Db	522	TGGAGATCAACACCAGCGGGGACITTCCTCACACAAGCGCTCAACCATATGTACAAATCC	581
Qy	181	GCACGAACCTCCAGCC	196
Db	582	GNACGAACCTNCAGCC	597

  

RESULT 8	
US-10-040-862-4407	
Sequence 4407, Application US/10040862	
Publication No. US20030078396A1	
GENERAL INFORMATION:	
APPLICANT: Gaiger, Alexander	
APPLICANT: Algate, Paul A.	
APPLICANT: Mannion, Jane	
APPLICANT: Retter, Marc	
APPLICANT: Corixa Corporation	
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and	
TITLE OF INVENTION: Hematological Malignancies	
FILE REFERENCE: 014058-013520US	
CURRENT APPLICATION NUMBER: US/10/040.862	
CURRENT FILING DATE: 2001-11-06	

RESULT 8  
US-10-040-862-4407  
: Sequence 4407, Application US/10040862  
: Publication No. US20030078396A1  
: GENERAL INFORMATION:  
: APPLICANT: Gaiger, Alexander  
: APPLICANT: Algate, Paul A.  
: APPLICANT: Mannion, Jane  
: APPLICANT: Retter, Marc  
: APPLICANT: Corixa Corporation  
: TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
: TITLE OF INVENTION: Hematological Malignancies  
: FILE REFERENCE: 014058-013520US  
: CURRENT APPLICATION NUMBER: US/10/040,862  
: CURRENT FILING DATE: 2001-11-06

;; PRIOR APPLICATION NUMBER: US 60/186,126  
;; PRIOR FILING DATE: 2000-03-01  
;; PRIOR APPLICATION NUMBER: US 60/190,479  
;; PRIOR FILING DATE: 2000-03-17  
;; PRIOR APPLICATION NUMBER: US 60/200,545  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: US 60/200,303  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: US 60/200,779  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: US 60/200,999  
;; PRIOR FILING DATE: 2000-05-01  
;; PRIOR APPLICATION NUMBER: US 60/202,084  
;; PRIOR FILING DATE: 2000-05-04  
;; PRIOR APPLICATION NUMBER: US 60/206,201  
;; PRIOR FILING DATE: 2000-05-22  
;; PRIOR APPLICATION NUMBER: US 60/218,950  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 60/222,903  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: US 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: US 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; PRIOR APPLICATION NUMBER: US 09/796,692  
;; PRIOR FILING DATE: 2001-03-01  
;; NUMBER OF SEQ ID NOS: 10467  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 4407  
;; LENGTH: 597  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (583)  
;; OTHER INFORMATION: n=A,T,C or G  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (592)  
;; OTHER INFORMATION: n=A,T,C or G  
US-10-040-862-4407

Query Match 64.7%; Score 194; DB 14; Length 597;  
Best Local Similarity 99.0%; Pred. No. 5.7e-55;  
Matches 194; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ACGAAATCCGACCCCTGTGAGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
Db 402 ACGAAATCCGACCCCTGTGAGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 461  
Qy 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 120  
Db 462 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 521  
Qy 121 TGGAGATCAACACCCAGCGGGACTTTCTCACACAAGCGCTCAACACATGTACAAACTCC 180  
Db 522 TGGAGATCAACACCCAGCGGGACTTTCTCACACAAGCGCTCAACACATGTACAAACTCC 581  
Qy 181 GCACGAACCTCCAGCC 196  
Db 582 GNACGAACCTCCAGCC 597

RESULT 9  
US-09-920-300A-682/c  
; Sequence 682, Application US/09920300A  
; Patent No. US20020136728A1  
; GENERAL INFORMATION:  
; APPLICANT: King, Gordon E.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

;; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
;; FILE REFERENCE: 210121.547  
;; CURRENT APPLICATION NUMBER: US/09/920,300A  
;; CURRENT FILING DATE: 2001-07-31  
;; NUMBER OF SEQ ID NOS: 1789  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 682  
;; LENGTH: 362  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-920-300A-682

Query Match 22.0%; Score 66; DB 10; Length 362;  
Best Local Similarity 100.0%; Pred. No. 4.6e-12;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 235 CCTGGTCAGCGGCTTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTCGTGG 294  
Db 362 CCTGGTCAGCGGCTTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTCGTGG 303  
Qy 295 TTCTGG 300  
Db 302 TTCTGG 297

RESULT 10  
US-09-998-598-2284/c  
; Sequence 2284, Application US/09998598  
; Patent No. US20020150922A1  
; GENERAL INFORMATION:  
; APPLICANT: Stolk, John A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Chenault, Ruth A.  
; APPLICANT: Meagher, Madeleine Joy  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.561  
; CURRENT APPLICATION NUMBER: US/09/998,598  
; CURRENT FILING DATE: 2001-11-16  
; NUMBER OF SEQ ID NOS: 2606  
; SOFTWARE: Corixa Invention Disclosure Database  
; SEQ ID NO 2284  
; LENGTH: 362  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-998-598-2284

Query Match 22.0%; Score 66; DB 10; Length 362;  
Best Local Similarity 100.0%; Pred. No. 4.6e-12;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 235 CCTGGTCAGCGGCTTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTCGTGG 294  
Db 362 CCTGGTCAGCGGCTTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTCGTGG 303  
Qy 295 TTCTGG 300  
Db 302 TTCTGG 297

RESULT 11  
US-10-099-926-682/c  
; Sequence 682, Application US/10099926  
; Publication No. US2003016064A1  
; GENERAL INFORMATION:  
; APPLICANT: King, Gordon E.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Secrist, Heather  
; APPLICANT: Jiang, Yugu  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.547C2

```
; CURRENT APPLICATION NUMBER: US/10/099,926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 682
; LENGTH: 362
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-099-926-682

Query Match          22.0%; Score 66; DB 12; Length 362;
Best Local Similarity 100.0%; Pred. No. 4.6e-12;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 CCTGGTCAGCGCGCTTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGG 294
Db 362 CCTGGTCAGCGCGCTTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGG 303

QY 295 TTCTGG 300
Db 302 TTCTGG 297

RESULT 12
US-10-033-528-682/c
; Sequence 682, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Serist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 682
; LENGTH: 362
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-682

Query Match          22.0%; Score 66; DB 13; Length 362;
Best Local Similarity 100.0%; Pred. No. 4.6e-12;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 CCTGGTCAGCGCGCTTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGG 294
Db 362 CCTGGTCAGCGCGCTTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGG 303

QY 295 TTCTGG 300
Db 302 TTCTGG 297

RESULT 13
US-10-027-632-151845/c
; Sequence 151845, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151845
; LENGTH: 820
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-151845

Query Match          13.1%; Score 39.2; DB 13; Length 820;
Best Local Similarity 60.2%; Pred. No. 0.006;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 23 GGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACACGCTTTGTGAGACA 82
Db 180 GAAGAAGTGTGAGCCTGGTTTCACCAAGTCTGGTGAAGGGCCCCCGCTACTTGTCTGCA 121

QY 83 GCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAA 130
Db 120 GGACGATGCCAGGCCAAGGTGGTCAACATCACCATCGTGCGCCCTCAA 73
```

```
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151845
; LENGTH: 820
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-151845

Query Match          13.1%; Score 39.2; DB 13; Length 820;
Best Local Similarity 60.2%; Pred. No. 0.006;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 23 GGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACACGCTTTGTGAGACA 82
Db 180 GAAGAAGTGTGAGCCTGGTTTCACCAAGTCTGGTGAAGGGCCCCCGCTACTTGTCTGCA 121

QY 83 GCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAA 130
Db 120 GGACGATGCCAGGCCAAGGTGGTCAACATCACCATCGTGCGCCCTCAA 73

RESULT 14
US-10-027-632-151846/c
; Sequence 151846, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151846
; LENGTH: 820
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-151846

Query Match          13.1%; Score 39.2; DB 13; Length 820;
Best Local Similarity 60.2%; Pred. No. 0.006;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 23 GGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACACGCTTTGTGAGACA 82
Db 180 GAAGAAGTGTGAGCCTGGTTTCACCAAGTCTGGTGAAGGGCCCCCGCTACTTGTCTGCA 121

QY 83 GCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAA 130
Db 120 GGACGATGCCAGGCCAAGGTGGTCAACATCACCATCGTGCGCCCTCAA 73
```

RESULT 15  
US-09-864-761-789  
; Sequence 789, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 789  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AP000215.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 10  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.6  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.9  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.4  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 11  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.8  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.2  
US-09-864-761-789

Query Match 10.9%; Score 32.8; DB 9; Length 468;

Best Local Similarity 54.0%; Pred. No. 0.71;  
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 82 AGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACCGGGGA 141  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 106 AGCAGAGCGCCCAAGGACAACTTGCAGAGGAGCTGCGTATCGAGCACTCACCGTGGGAGG 165  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 142 CTTTCTCTACACAAGCGCTCAACCCACATGTACAAACTCCGCACGACGAACTCCAGCCTCTCGG 201  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 166 CTTTCTCTAGGAGCGGAGCCAGCTCTCTGAGCTCCAGAGGACCTTGGCGCTGAGA 225  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 202 AGAG 205  
|||||  
Db 226 AGAG 229

Search completed: October 14, 2003, 17:17:33  
Job time : 148.333 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 ,Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 11:32:16 ; Search time 1272.67 Seconds  
(without alignments)  
5729.185 Million cell updates/sec

Title: US-09-854-124-7  
Perfect score: 300  
Sequence: 1 atcatgcttcagacacatc.....tggggattaaccttcagac 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	249.6	83.2	914	14	CA455226
2	245.4	81.8	709	10	BE561525
3	245.4	81.8	840	13	BU601226
4	244.8	81.6	624	10	BG284180
					CA455226 AGENCOURT
					BE561525 601345332
					BU601226 AGENCOURT
					BG284180 602408227

5	244.8	81.6	680	10	BE274253
6	244.8	81.6	736	10	BE312319
7	244.4	81.5	559	12	BM126492
8	244.4	81.5	601	12	BM043804
9	244.4	81.5	649	10	BE514071
10	244.4	81.5	659	13	BU729618
11	244.4	81.5	716	10	BE382866
12	244.4	81.5	801	12	BI196248
13	244.4	81.5	866	10	BF795157
14	244.4	81.5	876	10	BG181130
15	244.4	81.5	896	10	BE562088
16	244.4	81.5	914	10	BE547290
17	244.4	81.5	943	10	BF683514
18	244.4	81.5	1007	12	BM449472
19	244.4	81.5	1030	10	BF311745
20	244.4	81.5	1064	13	BQ277667
21	244.4	81.5	1071	13	BU184963
22	244.4	81.5	1182	11	AF125098
23	244.4	81.5	1201	9	AL582250
24	244.4	81.3	1201	9	AL582217
25	243.8	81.3	848	13	BU603101
26	243.4	81.1	833	13	BU597296
27	243	81.0	667	9	AU126087
28	242.8	80.9	488	14	CB160336
29	242.8	80.9	533	14	CB112523
30	242.8	80.9	696	9	AW249012
31	242.8	80.9	712	13	BU625683
32	242.8	80.9	714	10	BG104289
33	242.8	80.9	716	12	BG825252
34	242.8	80.9	730	10	BE296429
35	242.8	80.9	749	12	BG772776
36	242.8	80.9	790	10	BE795306
37	242.8	80.9	867	13	BQ233393
38	242.8	80.9	891	11	BC022839
39	242.8	80.9	921	10	BE799212
40	242.8	80.9	932	10	BE796384
41	242.8	80.9	1007	10	BE795838
42	242.8	80.9	1028	10	BE561044
43	241.8	80.6	909	13	BQ229290
44	241.2	80.4	628	10	BF035586
45	241.2	80.4	895	13	BU539659
					BE274253 601120527
					BE312319 601152780
					BM126492 1f06h06.y
					BM043804 603620658
					BE514071 601316376
					BU729618 01-E-CQ1-
					BE382866 601297755
					BI196248 602754709
					BF795157 602256302
					BG181130 602329264
					BE562088 601345038
					BE547290 601073715
					BF683514 602139714
					BM449472 AGENCOURT
					BF311745 601897244
					BQ277667 AGENCOURT
					BU184963 AGENCOURT
					AF125098 Homo sapi
					AL582250 AL582250
					AL582217 AL582217
					BU603101 AGENCOURT
					BU597296 AGENCOURT
					AU126087 K-EST0220
					CB160336 K-EST0154
					CB112523 K-EST0154
					AW249012 2821094.5
					BU625683 UI-H-FG1-
					BG104289 602310878
					BG825252 602747675
					BE296429 601176781
					BG772776 602720948
					BE795306 601586607
					BQ233393 AGENCOURT
					BC022839 Homo sapi
					BE799212 601592785
					BE796384 601592105
					BE795838 601590877
					BE561044 601344663
					BQ229290 AGENCOURT
					BF035586 601457517
					BU539659 AGENCOURT

#### ALIGNMENTS

RESULT 1  
CA455226  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CA455226  
AGENCOURT\_10714628 MAPcL Homo sapiens cdna clone IMAGE:6722756 5',  
mrna linear EST 12-NOV-2002  
AGENCOURT\_10714628 MAPcL Homo sapiens cdna clone IMAGE:6722756 5',  
mrna sequence.  
CA455226  
CA455226.1 GI:24905731  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 914)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cqabps-r@mail.nih.gov](mailto:cqabps-r@mail.nih.gov)  
Tissue Procurement: Kristi A. Egland, Ira Pastan  
CDNA Library Preparation: Invitrogen Corp  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM14285 row: p column: 20  
High quality sequence stop: 553.

FEATURES  
source

Location/Qualifiers  
1. .914  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6722756"  
/cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, HTERT-HME1  
, LNCap"  
/lab\_host="EMDH10B"  
/clone\_lib="MAPcL"  
/note="Vector: pCMV-SPORT6; Site\_1: EcoRV; Site\_2: Not I;  
Subtracted with brain, liver, lung, kidney and muscle.  
Directionally cloned. Priming method: oligo-dT. Average  
insert size: 1800 bp. Library amplification: 26,000 fold.  
Kristi A. Eglund, James J. Vincent, Robert Strausberg,  
Bungkok Lee & Ira Pastan: Discovery of new breast  
cancer genes encoding membrane and secreted proteins.  
Manuscript submitted."

BASE COUNT 246 a 233 c 246 g 188 t 1 others  
ORIGIN

Query Match 83.2%; Score 249.6; DB 14; Length 914;  
Best Local Similarity 94.1%; Pred. No. 1.9e-61;  
Matches 270; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 1 ATCATGCTTCAGACAACATCCGAGGAGGAGGAAATCCGGACCTGGTCAAGGATATGT 60  
|||||  
Db 361 ATCATGCTTCAGACAACATCCGAGGAGGAGGAAATCCGGACCTGGTCAAGGATATGT 420  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 120  
|||||  
Db 421 GGGACACTCGTATAGCCAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 480  
QY 121 CACATGCCAAGCTGATTAACCTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 180  
|||||  
Db 481 CACATGCCAAGCTGATTAACCTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 540  
QY 181 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTGGAAGACCTC 240  
|||||  
Db 541 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTGGAAGACCTC 600  
QY 241 AG--CTAGGACTTCTAATAAAGGCGTGTGACGCGCTGTGTTGGG 285  
|| |||||  
Db 601 AGTCTCAGGACTCTAATAAAGGCGTGTGACGCGCTGTGTTGGG 647

RESULT 2  
BE561525

LOCUS 601345332F1 NIH\_MGC\_8 709 bp mRNA linear EST 15-AUG-2000  
DEFINITION 601345332F1 NIH\_MGC\_8 Homo sapiens cDNA clone IMAGE:3678198 5',  
mRNA sequence.

ACCESSION BE561525  
VERSION BE561525.1 GI:9805245  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLC354 row: p column: 07  
High quality sequence start: 16  
High quality sequence stop: 706.

FEATURES  
source

Location/Qualifiers  
1. .709  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3678198"  
/tissue\_type="Burkitt lymphoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_8"  
/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGACGAG(S). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 185 a 178 c 203 g 142 t 1 others  
ORIGIN

Query Match 81.8%; Score 245.4; DB 10; Length 709;  
Best Local Similarity 94.8%; Pred. No. 2.8e-60;  
Matches 275; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACAACATCCGAGGAGGAGGAAATCCGGACCTGGTCAAGGATATGT 60  
|||||  
Db 393 ATCATGCTTCAGACAACATCCGAGGAGGAGGAAATCCGGACCTGGTCAAGGATATGT 452  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 120  
|||||  
Db 453 GGGACACTCGTATAGCCAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 512  
QY 121 CACATGCCAAGCTGATTAACCTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 180  
|||||  
Db 513 CACATGCCAAGCTGATTAACCTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 572  
QY 181 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTGGAAGACCTC 240  
|||||  
Db 573 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTGGAAGACCTC 632  
QY 241 AG--CTAGGACTTCTA-AAAAAGGCGTGTGACGCGCTGTGTTGGGAT 287  
|| |||||  
Db 633 AGTCTCAGGACTCTAGAGAAAGCGCTGTGACGCGGCTGTCTGGGGAT 682

RESULT 3  
BU601226

LOCUS BU601226 840 bp mRNA linear EST 20-SEP-2002  
DEFINITION AGENCOURT\_10029782 NIH\_MGC\_142 Homo sapiens cDNA clone  
IMAGE:6495109 5', mRNA sequence.

ACCESSION BU601226  
VERSION BU601226.1 GI:23252985  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

Tissue Procurement: NCI  
cDNA Library Preparation: Michael Brownstein Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2672 row: g column: 14  
High quality sequence stop: 496.  
Location/Qualifiers

```
source
1..840
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:6495109"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="PH10B (T1-phase-resistant)"
/clone_lib="NIH_MGC_142"
/note="Vector: pDNR-LIB; Site:1: SfiI (ggccattatggcc);
Site:2: SfiI (ggcgctcggcc); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
kidney - 2.3%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
used in cloning as follows:
5'-AAGCAGCTGTATCAACGACAGAGTGGCCATTAGCGCGGG-3' and
5'-ATTCTAGAGCGGCGGCGGACATG-DT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the >0.5 kb
size fraction (other fractions present in NIH_MGC_141).
Library created in the laboratory of M. Brownstein (NIMH,
NIH). Note: this is a NIH_MGC Library."
BASE COUNT 217 a 215 c 221 g 185 t 2 others
ORIGIN

Query Match 81.8%; Score 245.4; DB 13; Length 840;
Best Local Similarity 93.0%; Pred. No. 3e-60;
Matches 267; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

QY 1 ATCATGCTTCAGACACATCCGAGGCGAGAGCAATCCGGACCCCTGGTCAAGGATATGT 60
DB 340 ATTCATGCTTCAGACACATCCGAGGCGAGAGCAATCCGGACCCCTGGTCAAGGATATGT 399
QY 61 GGCACACTCGTATAGCCAACTCCGAGTGTCTGTCGACAGCTTGTGAGACACGAGGAG 120
DB 400 GGCACACTCGTATAGCCAACTCCGAGTGTCTGTCGACAGCTTGTGAGACACGAGGAG 459
QY 121 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 180
DB 460 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 519
QY 181 CACAAGCGCTCAACACATGTCACAACTCCGACGACGAACTCCAGCCTCTGGAAGACCTC 240
DB 520 CACAAGCGCTCAACACATGTCACAACTCCGACGACGAACTCCAGCCTCTGGAAGACCTC 579
QY 241 AG-CTAGGACTTCTAAAAAGCGCTGGTGTGACGCGCTTGGTTGGGG 285
DB 580 AGTCTCAGACTTCTANAGAAGCGCTGGTGTGACGCGCTTGGTTGGGG 626

RESULT 4
BG284180 624 bp mrna linear EST 21-FEB-2001
LOCUS 602408227F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4520430 5',
DEFINITION mRNA sequence.
ACCESSION BG284180
VERSION BG284180.1 GI:13034868
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 624)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10418 row: e column: 07
High quality sequence start: 4
High quality sequence stop: 621.
Location/Qualifiers
1..624
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:4520430"
/tissue_type="adenocarcinoma, cell line"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NIH_MGC_91"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 163 a 168 c 166 g 126 t 1 others
ORIGIN

Query Match 81.8%; Score 244.8; DB 10; Length 624;
Best Local Similarity 94.8%; Pred. No. 4e-60;
Matches 275; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACACATCCGAGGCGAGAGCAATCCGGACCCCTGGTCAAGGATATGT 60
DB 333 ATCATGCTTCAGACACATCCGAGGCGAGAGCAATCCGGACCCCTGGTCAAGGATATGT 392
QY 61 GGCACACTCGTATAGCCAACTCCGAGTGTCTGTCGACAGCTTGTGAGACACGAGGAG 120
DB 393 GGCACACTCGTATAGCCAACTCCGAGTGTCTGTCGACAGCTTGTGAGACACGAGGAG 452
QY 121 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 180
DB 453 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 512
QY 181 CACAAGCGCTCAACACATGTCACAACTCCGACGACGAACTCCAGCCTCTGGAAGACCTC 240
DB 513 CACAAGCGCTCAACACATGTCACAACTCCGACGACGAACTCCAGCCTCTGGAAGACTC 572
QY 241 AG-CTAGGACTTCTA-AAAAAGCGCTGGTGTGACGCGCTTGGTTGGGGAT 287
DB 573 AGTCTCAGGACTTCTAGAGAAGCGCTGGTGTGACGCGCTGGTGGGGAT 622

RESULT 5
BE274253 680 bp mrna linear EST 13-JUL-2000
LOCUS 601120527F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2967175 5',
DEFINITION mRNA sequence.
ACCESSION BE274253
VERSION BE274253.1 GI:9149193
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 680)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
```

Plate: LLCW71 row: f column: 08  
 High quality sequence stop: 651.  
 FEATURES source  
 Location/Qualifiers  
 1. .880

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2967175"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_20"  
 /note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 174 a 178 C 186 g 142 t  
 ORIGIN

Query Match 81.6%; Score 244.8; DB 10; Length 680;  
 Best Local Similarity 94.8%; Pred. No. 4.1e-60;  
 Matches 275; Conservative 0; Mismatches 12; Indels 3; Gaps 2;  
 QY 1 ATCATGCTTCAGACAAACATCCGAGGCGAGAGAAATCCGACCTGTGTCAGGATATGT 60  
 Db 318 ATCATGCTTCAGACAAACATCCGAGGCGAGAGAAATCCGACCTGTGTCAGGATATGT 60  
 QY 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 120  
 Db 378 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 437  
 QY 121 CACATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCGCGGACTTTCCCTCA 180  
 Db 438 CACATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCGCGGACTTTCCCTCA 497  
 QY 181 CACAGCGCTCAACCATGATGACAACTCCGACGAGAACTCCAGCTCTGGAAGACCTC 240  
 Db 498 CACAGCGCTCAACCATGATGACAACTCCGACGAGAACTCCAGCTCTGGAAGACCTC 557  
 QY 241 AG--CTAGGACTTCTA-AAAAAGCGCTGGTGCAGCGCTTGTGGGGAT 287  
 Db 558 AGTCTCAGGACTTCTAGAGAAGCGCTGGTGCAGCGCTGCTGGGGAT 607

RESULT 6  
 BE312319  
 LOCUS BE312319 736 bp mRNA linear EST 26-OCT-2000  
 DEFINITION 601152780F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:3508866 5', mRNA sequence.  
 ACCESSION BE312319  
 VERSION BE312319.1 GI:9130832  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 736)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLCM190 row: p column: 19  
 High quality sequence stop: 645.

FEATURES source  
 Location/Qualifiers  
 1. .736

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3508866"  
 /tissue\_type="neuroblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_19"  
 /note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 Note: this is a NIH\_MGC library.  
 BASE COUNT 182 a 196 c 208 g 150 t  
 ORIGIN

Query Match 81.6%; Score 244.8; DB 10; Length 736;  
 Best Local Similarity 94.8%; Pred. No. 4.3e-60;  
 Matches 275; Conservative 0; Mismatches 12; Indels 3; Gaps 2;  
 QY 1 ATCATGCTTCAGACAAACATCCGAGGCGAGAGAAATCCGACCTGTGTCAGGATATGT 60  
 Db 367 ATCATGCTTCAGACAAACATCCGAGGCGAGAGAAATCCGACCTGTGTCAGGATATGT 426  
 QY 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 120  
 Db 427 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGG 486  
 QY 121 CACATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCGCGGACTTTCCCTCA 180  
 Db 487 CACATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCGCGGACTTTCCCTCA 546  
 QY 181 CACAGCGCTCAACCATGATGACAACTCCGACGAGAACTCCAGCTCTGGAAGACCTC 240  
 Db 547 CACAGCGCTCAACCATGATGACAACTCCGACGAGAACTCCAGCTCTGGAAGACCTC 606  
 QY 241 AG--CTAGGACTTCTA-AAAAAGCGCTGGTGCAGCGCTTGTGGGGAT 287  
 Db 607 AGTCTCAGGACTTCTAGAGAAGCGCTGGTGCAGCGCTGCTGGGGAT 656

RESULT 7  
 BM126492  
 LOCUS BM126492 559 bp mRNA linear EST 12-MAR-2002  
 DEFINITION if06h06.yl Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA clone IMAGE:5675770 5', similar to TR:Q9Y248 Q9Y248 HSPC037 PROTEIN. ;, mRNA sequence.  
 ACCESSION BM126492  
 VERSION BM126492.1 GI:17121044  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 559)  
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T., Jackson, Y., and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished  
 Other ESTs: if06h06.xl  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812  
Fax: 617-495-8557

Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Juliana Brown  
(brownefas.harvard.edu) This sequence now available from the IMAGE  
consortium, for clone orders contact: info@image.llnl.gov  
High quality sequence stop: 437.

#### FEATURES

Location/Qualifiers  
1..559  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5675770"  
/sex="Both"  
/tissue\_type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="Melton Normalized Human Islet 4 M4-HIS 1"  
/note="Organ: Pancreas; Vector: pSPORT1; Site\_1: Not 1;  
Site\_2: Sal 1; Starting library constructed using  
SuperScript plasmid library kit (Life Technologies). cDNA  
made by oligo-dT priming. Size-selected by column  
fractionation; average insert size 1.08 kb. Library was  
amplified once on solid support and plasmid DNA from  
library was prepared. The library DNA was normalized by  
method #4 from Bonaldo, Lennon, and Soares 1996 Genome  
Research 6:791-806; 0.5 microgram single-stranded library  
plasmid DNA was mixed with 5 micrograms PCR product  
representing library inserts and hybridized to an EcoT of  
20. Single-stranded (unhybridized) plasmids were isolated  
by hydroxyapatite chromatography and used to make this  
library."

BASE COUNT 157 a 131 c 142 g 129 t  
ORIGIN

Query Match 81.5%; Score 244.4; DB 12; Length 559;  
Best Local Similarity 95.1%; Pred. No. 4.9e-60;  
Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACACATCCCGAGGACAGCAATCCGGACCCCTGGTCAAGGATATGT 60  
Db |||||  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 120  
Db |||||  
QY 148 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 207  
QY 121 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCAGCGGACTTTCTCTCA 180  
Db |||||  
QY 208 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCAGCGGACTTTCTCTCA 267  
QY 181 CACAAGCGCTACACCATGTACAACTCCGACAGCACTCCAGCTCTGGAAAGACCTC 240  
Db |||||  
QY 268 CACAAGCGCTACACCATGTACAACTCCGACAGCACTCCAGCTCTGGAGAGTACTC 327.  
QY 241 AG--CTPAGACTTCTA-AAAAAGGCGCTGGTGACGCGCTTGGTTGGG 285  
Db |||||  
QY 328 AGTCTCAGGACTTCTAGAGAAAGGCGCTGGTGACGCGGCTTGGTGGGG 375

#### RESULT 8

BM043804  
LOCUS 603620658F1 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:5446215 5',  
DEFINITION mRNA sequence.  
ACCESSION BM043804  
VERSION BM043804.1 GI:16773071  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

#### REFERENCE

1 (bases 1 to 601)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1928 row: c column: 16  
High quality sequence stop: 598.

#### FEATURES

Location/Qualifiers  
1..601  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5446215"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_40"  
/note="Organ: prostate; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

BASE COUNT 161 a 158 c 159 g 123 t  
ORIGIN

Query Match 81.5%; Score 244.4; DB 12; Length 601;  
Best Local Similarity 95.1%; Pred. No. 5.1e-60;  
Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACACATCCCGAGGACAGCAATCCGGACCCCTGGTCAAGGATATGT 60  
Db |||||  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 120  
Db |||||  
QY 366 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 425  
QY 121 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCAGCGGACTTTCTCTCA 180  
Db |||||  
QY 426 CACATGCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCAGCGGACTTTCTCTCA 485  
QY 181 CACAAGCGCTACACCATGTACAACTCCGACAGCACTCCAGCTCTGGAAAGACCTC 240  
Db |||||  
QY 486 CACAAGCGCTACACCATGTACAACTCCGACAGCACTCCAGCTCTGGAGAGTACTC 545  
QY 241 AG--CTPAGACTTCTA-AAAAAGGCGCTGGTGACGCGCTTGGTTGGG 285  
Db |||||  
QY 546 AGTCTCAGGACTTCTAGAGAAAGGCGCTGGTGACGCGGCTTGGTGGGG 593

#### RESULT 9

BE514071  
LOCUS 601316576F1 NIH\_MGC\_8 Homo sapiens cDNA clone IMAGE:3635010 5',  
DEFINITION mRNA sequence.  
ACCESSION BE514071  
VERSION BE514071.1 GI:9721283  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: LCM330 row: h column: 19
High quality sequence stop: 646.
Location/Qualifiers
1..649
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3635010"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_8"
/notes="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 168 a 173 c 176 g 132 t

Query Match 81.5%; Score 244.4; DB 10; Length 649;  
Best Local Similarity 95.1%; Pred. No. 5.3e-60;  
Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

```

QY 1 ATCATGCTTCAGACAACATCCGGAAGGCAGACGAAATCCGACCTTGTCAAGGATATGT 60
DB |||||||
QY 330 ATCATGCTTCAGACAACATCCGGAAGGCAGACGAAATCCGACCTTGTCAAGGATATGT 389
DB |||||||
QY 61 GGGACACTCGTATAGCCAAATCCGAGTGCTGCTGACAGCTTTGTGAGACAGCAGGAGG 120
DB |||||||
QY 390 GGGACACTCGTATAGCCAAATCCGAGTGCTGCTGACAGCTTTGTGAGACAGCAGGAGG 449
DB |||||||
QY 121 CACATGCCAAGCTGGATAACTTACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 180
DB |||||||
QY 450 CACATGCCAAGCTGGATAACTTACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 509
DB |||||||
QY 181 CACAAGCGCTCAACCATGTACAAATCCGACGACCTCCAGCCCTCGGAAGACCTC 240
DB |||||||
QY 510 CACAAGCGCTCAACCATGTACAAATCCGACGACCTCCAGCCCTCGGAAGACCTC 240
DB |||||||
QY 241 AG--CTAGGACTTCTA-AAAAAGCCCTGGTGACCGCTTGGTGGGG 285
DB |||||||
QY 570 AGTCTCAGGACTTCTAGAGAAGCCCTGGTGACCGCGCTTGGTGGGG 617
DB |||||||

RESULT 10
BU729618/c
LOCUS
DEFINITION BU729618 659 bp mRNA linear EST 09-OCT-2002
UI-E-CQ1-ay-b-18-0-UI.s1 UI-E-CQ1 Homo sapiens cDNA clone
UI-E-CQ1-ay-b-18-0-UI.3', mRNA sequence.
ACCESSION BU729618.1 GI:23652681
VERSION BU729618
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 659)
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
```

Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
Seq primer: M13 FORWARD  
POLYA=Yes.

Location/Qualifiers  
1..659  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-CQ1-ay-b-18-0-UI"  
/tissue\_type="optic nerve"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-CQ1"  
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-CQ1 is a normalized cDNA library containing the following tissue(s): optic nerve. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dr primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag (dN18 tail. The sequence tag for this library is CCATTAAGTG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
TAG LIB-UI-E-CQ1  
TAG\_TISSUE=human optic nerve  
TAG\_SEQ=CCATTAAGTG"

BASE COUNT 149 a 159 c 151 g 200 t

ORIGIN

Query Match 81.5%; Score 244.4; DB 13; Length 659;  
Best Local Similarity 95.1%; Pred. No. 5.3e-60;  
Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

```

QY 1 ATCATGCTTCAGACAACATCCGGAAGGCAGACGAAATCCGACCTTGTCAAGGATATGT 60
DB |||||||
QY 539 ATCATGCTTCAGACAACATCCGGAAGGCAGACGAAATCCGACCTTGTCAAGGATATGT 480
DB |||||||
QY 61 GGGACACTCGTATAGCCAAATCCGAGTGCTGCTGACAGCTTTGTGAGACAGCAGGAGG 120
DB |||||||
QY 479 GGGACACTCGTATAGCCAAATCCGAGTGCTGCTGACAGCTTTGTGAGACAGCAGGAGG 420
DB |||||||
QY 121 CACATGCCAAGCTGGATAACTTACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 180
DB |||||||
QY 419 CACATGCCAAGCTGGATAACTTACCTTGATGGAGATCAACACAGCGGGACTTTCCTCA 360
DB |||||||
QY 181 CACAAGCGCTCAACCATGTACAAATCCGACGACCTCCAGCCCTCGGAAGACCTC 240
DB |||||||
QY 359 CACAAGCGCTCAACCATGTACAAATCCGACGACCTCCAGCCCTCGGAAGACCTC 300
DB |||||||
QY 241 AG--CTAGGACTTCTA-AAAAAGCCCTGGTGACCGCGCTTGGTGGGG 285
DB |||||||
```

Db 299 AGTCTCAGGACTTCTAGAGAAAGGCTTGGTGCAGCGGCTTGTCTGGGG 252

## RESULT 11

BE382866

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE382866 716 bp mRNA linear EST 21-JUL-2000

60129775F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:3627852 5',

mRNA sequence.

BE382866

BE382866.1 GI:9328231

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 716)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM311 row: n column: 13

High quality sequence stop: 651.

Location/Qualifiers

1. .716

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3627852"

/tissue\_type="neuroblastoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_19"

/notes="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site:2:

ECORI; cDNA made by oligo-dt priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

BASE COUNT 185 a 179 c 195 g 157 t

ORIGIN

Query Match

Best Local Similarity 81.5%; Score 244.4; DB 10; Length 716;

Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACACATCCCGAAGCGAGACGAAATCCGGACCTTGGTCAAGGATATGT 60

DB 307 ATCATGCTTCAGACACATCCCGAAGCGAGACGAAATCCGGACCTTGGTCAAGGATATGT 366

QY 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGACAGCAGCAGGAGG 120

DB 367 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGACAGCAGCAGGAGG 426

QY 121 CACATGCCAAGCTGGATAACTTTGACCTTGTATGGAGATCAACACCGAGCGGACTTTCTCTCA 180

DB 427 CACATGCCAAGCTGGATAACTTTGACCTTGTATGGAGATCAACACCGAGCGGACTTTCTCTCA 486

QY 181 CACAAGCGCTCAACACATGTACAACTCCGACGAGAACTCCAGCTTCCAGCTTGGAAAGACCTC 240

DB 487 CACAAGCGCTCAACACATGTACAACTCCGACGAGAACTCCAGCTTCCAGCTTGGAAAGACCTC 240

QY 241 AG--CTAGACTTCTA-AAAAAGGCGCTGGTGCAGCGGCTTGTCTGGGG 285

DB 547 AGTCTCAGGACTTCTAGAGAAAGGCGCTGGTGCAGCGGCTTGTCTGGGG 594

## RESULT 12

BI196248

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI196248 801 bp mRNA linear EST 10-JUL-2001

602794709F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4890147 5',

mRNA sequence.

BI196248

BI196248.1 GI:14651268

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 801)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM1767 row: b column: 04

High quality sequence stop: 781.

Location/Qualifiers

1. .801

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4890147"

/tissue\_type="neuroblastoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_19"

/notes="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site:2:

ECORI; cDNA made by oligo-dt priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

BASE COUNT 208 a 196 c 216 g 181 t

ORIGIN

Query Match

Best Local Similarity 81.5%; Score 244.4; DB 12; Length 801;

Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACACATCCCGAAGCGAGACGAAATCCGGACCTTGGTCAAGGATATGT 60

DB 326 ATCATGCTTCAGACACATCCCGAAGCGAGACGAAATCCGGACCTTGGTCAAGGATATGT 385

QY 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGACAGCAGCAGGAGG 120

DB 386 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGACAGCAGCAGGAGG 445

QY 121 CACATGCCAAGCTGGATAACTTTGACCTTGTATGGAGATCAACACCGAGCGGACTTTCTCTCA 180

DB 446 CACATGCCAAGCTGGATAACTTTGACCTTGTATGGAGATCAACACCGAGCGGACTTTCTCTCA 505

QY 181 CACAAGCGCTCAACACATGTACAACTCCGACGAGAACTCCAGCTTCCAGCTTGGAAAGACCTC 240

DB 506 CACAAGCGCTCAACACATGTACAACTCCGACGAGAACTCCAGCTTCCAGCTTGGAAAGACCTC 240

QY 241 AG--CTAGACTTCTA-AAAAAGGCGCTGGTGCAGCGGCTTGTCTGGGG 285

DB 566 AGTCTCAGGACTTCTAGAGAAAGGCGCTGGTGCAGCGGCTTGTCTGGGG 613



```

RESULT 13
BF795157
LOCUS      BF795157
DEFINITION 602256302F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4339537 5',
            mRNA sequence.
ACCESSION  BF795157
VERSION     BF795157
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 866)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9950 row: d column: 02
            High quality sequence stop: 740.
FEATURES    source
            1..866
                location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4339537"
                /tissue_type="lymphoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_85"
                /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.867 Kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT  220 a 221 c 236 g 189 t
            Query Match      81.5%; Score 244.4; DB 10; Length 866;
            Best Local Similarity 95.1%; Pred. No. 6e-60;
            Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

Qy 1 ATCATGCTTCAGACAACATCCGAGGCAGACGAAATCCGGACCTGGTCAAGGATATGT 60
Db 367 ATCATGCTTCAGACAACATCCGAGGCAGACGAAATCCGGACCTGGTCAAGGATATGT 426

Qy 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 120
Db 427 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 486

Qy 121 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACAGCGGGACTTTCTCTCA 180
Db 487 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACAGCGGGACTTTCTCTCA 546

Qy 181 CACAAGCGCTCAACCATGTACAACTCCGACGACCACTCCAGCCCTCTGGAAGACCTC 240
Db 547 CACAAGCGCTCAACCATGTACAACTCCGACGACCACTCCAGCCCTCTGGAAGACCTC 606

Qy 241 AG--CTAGGACTTCTA-AAAAAGCGCTGGTGCAGCGCTTGGTTGGGG 285
Db 607 AGTCTCAGGACTTCTAGAGAAGGCCCTGGTGCAGCGCTTGGTTGGGG 654

RESULT 14
BF795157
LOCUS      BF795157
DEFINITION 602329264F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4430823 5',
            mRNA sequence.
ACCESSION  BF795157
VERSION     BF795157
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 876)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10184 row: o column: 16
            High quality sequence stop: 674.
FEATURES    source
            1..876
                location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4430823"
                /tissue_type="adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_91"
                /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.4 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT  246 a 217 c 214 g 199 t
            Query Match      81.5%; Score 244.4; DB 10; Length 876;
            Best Local Similarity 95.1%; Pred. No. 6e-60;
            Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

Qy 1 ATCATGCTTCAGACAACATCCGAGGCAGACGAAATCCGGACCTGGTCAAGGATATGT 60
Db 85 ATCATGCTTCAGACAACATCCGAGGCAGACGAAATCCGGACCTGGTCAAGGATATGT 144

Qy 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 120
Db 145 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 204

Qy 121 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACAGCGGGACTTTCTCTCA 180
Db 205 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACAGCGGGACTTTCTCTCA 264

Qy 181 CACAAGCGCTCAACCATGTACAACTCCGACGACCACTCCAGCCCTCTGGAAGACCTC 240
Db 265 CACAAGCGCTCAACCATGTACAACTCCGACGACCACTCCAGCCCTCTGGAAGACCTC 324

Qy 241 AG--CTAGGACTTCTA-AAAAAGCGCTGGTGCAGCGCTTGGTTGGGG 285
Db 325 AGTCTCAGGACTTCTAGAGAAGGCCCTGGTGCAGCGCTTGGTTGGGG 372

RESULT 15
BF795157
LOCUS      BF795157
DEFINITION 601345038F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3678080 5',
            mRNA sequence.
ACCESSION  BF795157
VERSION     BF795157
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 896)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10184 row: o column: 16
            High quality sequence stop: 674.
FEATURES    source
            1..896
                location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4430823"
                /tissue_type="adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_91"
                /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.4 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT  246 a 217 c 214 g 199 t
            Query Match      81.5%; Score 244.4; DB 10; Length 876;
            Best Local Similarity 95.1%; Pred. No. 6e-60;
            Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

Qy 1 ATCATGCTTCAGACAACATCCGAGGCAGACGAAATCCGGACCTGGTCAAGGATATGT 60
Db 85 ATCATGCTTCAGACAACATCCGAGGCAGACGAAATCCGGACCTGGTCAAGGATATGT 144

Qy 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 120
Db 145 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 204

Qy 121 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACAGCGGGACTTTCTCTCA 180
Db 205 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACAGCGGGACTTTCTCTCA 264

Qy 181 CACAAGCGCTCAACCATGTACAACTCCGACGACCACTCCAGCCCTCTGGAAGACCTC 240
Db 265 CACAAGCGCTCAACCATGTACAACTCCGACGACCACTCCAGCCCTCTGGAAGACCTC 324

Qy 241 AG--CTAGGACTTCTA-AAAAAGCGCTGGTGCAGCGCTTGGTTGGGG 285
Db 325 AGTCTCAGGACTTCTAGAGAAGGCCCTGGTGCAGCGCTTGGTTGGGG 372

```



SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 896)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LCM354 row: k column: 09  
High quality sequence start: 24  
High quality sequence stop: 840.  
Location/Qualifiers  
1..896  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3678080"  
/tissue\_type="Burkitt lymphoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_8"  
/notes="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 218 a 230 c 251 g 197 t  
ORIGIN  
Query Match 81.5%; Score 244.4; DB 10; Length 896;  
Best Local Similarity 95.1%; Pred. No. 6.1e-60;  
Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;  
QY 1 ATCATGCTTCAGACACATCCCGAGGCGAGACGAAATCCGGACCCCTGGTCAAGGATATGT 60  
Db |||||  
408 ATCATGCTTCAGACACATCCCGAGGCGAGACGAAATCCGGACCCCTGGTCAAGGATATGT 467  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACACGAGG 120  
Db |||||  
468 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACACGAGG 527  
QY 121 CACATGCCAAGCTGGGATACTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 180  
Db |||||  
528 CACATGCCAAGCTGGGATACTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 587  
QY 181 CACAAGCGCTCAACACCATGTACAACTCCGACACGAACTCCAGCCCTCTGGAAGACCTC 240  
Db |||||  
588 CACAAGCGCTCAACACCATGTACAACTCCGACACGAACTCCAGCCCTCTGGAAGTACTC 647  
QY 241 AG--CTAGGACTTCTA-AAAAAGCGCTGGTGCAGCGCTTGGTGGG 285  
Db |||||  
648 AGTCTCAGGACTTCTAGAGAAAGGCGCTGGTGCAGCGCTTGGTGGG 695

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 14:17:11 ; Search time 143.333 Seconds  
(without alignments)  
5431.000 Million cell updates/sec

Title: US-09-854-124-7  
Perfect score: 300  
Sequence: 1 atcagtttcagacaacatc.....tggggatttaacccttcagac 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/ECT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/ECTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	300	100.0	300	9	US-09-854-124-7
2	244.4	81.5	1210	9	Sequence 7, Appl1
3	237.2	79.1	484	11	Sequence 214, App
4	224	74.7	597	10	Sequence 19338, A
5	224	74.7	597	10	Sequence 4407, Ap
6	214.4	71.5	300	9	Sequence 4407, Ap
7	199.2	66.4	382	14	Sequence 5, Appl1
8	199.2	66.4	382	14	Sequence 9460, Ap
9	39.2	13.1	820	13	Sequence 9460, Ap
10	39.2	13.1	820	13	Sequence 151845, Ap
11	35	11.7	481	11	Sequence 151846, Ap
12	33.4	11.1	476	11	Sequence 870, App
13	32.6	10.9	39061	12	Sequence 2536, Ap
14	32.6	10.9	39061	12	Sequence 148, App
15	31.8	10.6	468	9	Sequence 254, App
16	31.8	10.6	737	9	Sequence 789, App
					Sequence 17572, A

17	31.8	10.6	814	13	US-10-027-632-156310
18	31.6	10.5	650	13	Sequence 156310, Sequence 246523, Sequence 1387, Ap
19	31.4	10.5	517	10	US-09-998-598-1387
20	31.4	10.5	596	13	Sequence 50355, A
21	31.4	10.5	596	13	Sequence 50355, A
22	31.4	10.5	596	13	Sequence 50356, A
23	31.4	10.5	596	13	Sequence 50356, A
24	31.2	10.4	546	10	Sequence 2478, Ap
25	31.2	10.4	3309400	10	Sequence 1, Appl1
26	31	10.3	408	10	Sequence 14278, A
27	31	10.3	434	10	Sequence 12424, A
28	31	10.3	717	14	Sequence 43, Appl1
29	31	10.3	1623	14	Sequence 33, Appl1
30	31	10.3	1827	11	Sequence 608, App
31	31	10.3	1827	14	Sequence 608, App
32	31	10.3	2439	14	Sequence 3, Appl1
33	31	10.3	2439	14	Sequence 5, Appl1
34	31	10.3	3171	14	Sequence 21, Appl1
35	31	10.3	4833	14	Sequence 31, Appl1
36	31	10.3	4944	13	Sequence 51, Appl1
37	31	10.3	6811	11	Sequence 51, Appl1
38	31	10.3	8999	13	Sequence 2, Appl1
39	30.8	10.3	487	9	Sequence 48, Appl1
40	30.8	10.3	558	9	Sequence 291, App
41	30.8	10.3	811	13	Sequence 7880, Ap
42	30.8	10.3	1371	14	Sequence 140232, Sequence 5598, Ap
43	30.8	10.3	3051	13	Sequence 111, App
44	30.8	10.3	9025608	14	Sequence 1, Appl1
45	30.4	10.1	476	11	Sequence 19581, A

## ALIGNMENTS

## RESULT 1

US-09-854-124-7  
; Sequence 7, Application US/09854124  
; Patent No. US20020076735A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Escobedo, Jaime  
; APPLICANT: Innis, Michael A.  
; APPLICANT: Garcia, Pablo Dominiguez  
; APPLICANT: Sudduth-Klinger, Julie  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Giese, Klaus  
; APPLICANT: Randazzo, Filippo  
; APPLICANT: Kennedy, Giulia C.  
; APPLICANT: Pot, David  
; APPLICANT: Kassam, Altaf  
; APPLICANT: Lamson, George  
; APPLICANT: Drmanac, Radoje  
; APPLICANT: Crkvenjakov, Radomir  
; APPLICANT: Dickson, Mark  
; APPLICANT: Drmanac, Shezana  
; APPLICANT: Labat, Ivan  
; APPLICANT: Leshkowitz, Dena  
; APPLICANT: Kita, David  
; APPLICANT: Jones, Lee William  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Stache-Crain, Birgit  
; TITLE OF INVENTION: Diagnostic and Therapeutic Methods Using Molecules Differentially Expressed in Cancer Cells  
; FILE REFERENCE: 2300-1490  
; CURRENT APPLICATION NUMBER: US/09/854,124  
; CURRENT FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: 09/400,947  
; PRIOR FILING DATE: 1999-09-22  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 300  
; TYPE: DNA

RESULT 4  
US-09-796-692-4407  
; Sequence 4407, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4407  
; LENGTH: 597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (583)  
; NAME/KEY: unsure  
; LOCATION: (592)  
; OTHER INFORMATION: n=A,T,C or G  
; OTHER INFORMATION: n=A,T,C or G  
US-09-796-692-4407

Query Match 74.7%; Score 224; DB 10; Length 597;

Best Local Similarity 99.1%; Pred. No. 4.7e-67;

Matches 224; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCATGCTTCAGACAAACATCCGAGGCGAGAGAGAAATCCGGACCCCTGGTCAAGGATATGT 60  
DB 372 ATCATGCTTCAGACAAACATCCGAGGCGAGAGAGAAATCCGGACCCCTGGTCAAGGATATGT 431  
QY 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTGTGACACAGCAGGAGG 120  
DB 432 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTGTGACACAGCAGGAGG 491  
QY 121 CATATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCGGGGACTTTCCCTCA 180  
DB 492 CATATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCGGGGACTTTCCCTCA 551  
QY 181 CACAGCGCTCAACACATGTACAACTCCGACGAACTCCAGCC 226  
DB 552 CACAAGCGCTCAACACATGTACAACTCCGACGAACTCCAGCC 597

#### RESULT 5

US-10-040-862-4407

; Sequence 4407, Application US/10040862

; Publication No. US20030078396A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; APPLICANT: Retter, Marc

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy

; FILE REFERENCE: 014058-01352005

; CURRENT APPLICATION NUMBER: US/10/040,862

; CURRENT FILING DATE: 2001-11-06

; PRIOR APPLICATION NUMBER: US 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: US 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: US 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: US 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: US 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-04-28

```
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: Diagnostic and Therapeutic Methods Using
; TITLE OF INVENTION: Molecules Differentially Expressed in Cancer Cells
; FILE REFERENCE: 2300-1490
; CURRENT APPLICATION NUMBER: US/09/854,124
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/400,947
; PRIOR FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-854-124-5

Query Match          71.5%; Score 214.4; DB 9; Length 300;
Best Local Similarity 94.6%; Pred. No. 7.5e-64;
Matches 244; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 31 ACGAATCCGACCCCTGTCAGGATATGTGGACATCTGATAGCCAAACTCCGAGTGT 90
Db 1 ACGAATCCGACCCCTGTCAGGATATGTGGACATCTGATAGCCAAACTCCGAGTGT 60

QY 91 CTGCTGACAGCTTTGTGAGACAGAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 150
Db 61 CTGCTGACAGCTTTGTGAGACAGAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 120

QY 151 TGGAGATCAACACAGCGGGGACTTCTCCACACAGGCTCAACACATGTACAACTCC 210
Db 121 TGGAGATCAACACAGCGGGGACTTCTCCACACAGGCTCAACACATGTACAACTCC 180

QY 211 GCAGCAACTCCAGCCTCTGAAAGACCTCAG--CTAGGACTTCTA-AAAAAGGCCTGGT 267
Db 181 GCAGCAACTCCAGCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCCTGGT 240

QY 268 GCAGCCGCTTGGTTGGGG 285
Db 241 GCAGCCGCTTGGTTGGGG 258

RESULT 7
US-09-796-692-9460/c
; Sequence 9460, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084

; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9460
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (72)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (80)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (89)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (122)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (129)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (131)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (155)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (173)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (222)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (235)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (236)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (237)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (245)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (251)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (261)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (305)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
```



```
FEATURE: unsure
NAME/KEY: (237)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (245)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (251)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (261)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (305)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (312)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (321)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (334)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (352)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (354)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (373)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: (375)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-9460
```

```
Query Match 66.4%; Score 199.2; DB 14; Length 382;
Best Local Similarity 91.4%; Pred. No. 1.5e-58;
Matches 223; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

Qy 1 ATCATGCTCAGACAACTCCGAGGACGACGAAATCCGGACCCTGGTCAAGGATATGT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 ATCATNCTCAGACANCACTCCGAGGACGACGAAATCCGGACCCTGGTCAAGGATATGT 207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGTTTGTGAGACA-GCAGGAG 119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGNTGACAGTTTGTGAGACNCGCAGGAG 147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 120 GCATATCCCAAGCTGGATACTTGACCTTGATGGAGATCAACACCGGGACTTTCCCTC 179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 146 GCATATCCCAAGCTGNAACCTTGNCTTGTGAGATCAACACCGGGNCTTTCNTC 87
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 180 ACACA-AGCGCTACACACATGTACAACTCCGACGACAACTCCAGCCTCTGGAAGACC 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86 ACACATNGCGCTCANCACATGTACAACTCCGACGACAACTCCGACGACCTCTGGAGATAC 27
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 239 TCAG 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 26 TCAG 23
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 9
US-10-027-632-151845/c
; Sequence 151845, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151845
; LENGTH: 820
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-151845

Query Match 13.1%; Score 39.2; DB 13; Length 820;
Best Local Similarity 60.2%; Pred. No. 0.0037;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 53 GGATATGTGGACACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACA 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180 GAACAAGTGTGAGCCTGTTTCAACCAAGTCTGGTGAAGGGCCCGCTACTTGTCTGCA 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 113 GCAGGAGGCACATGCCAAGTGGATAACTTGACCTTGATGAGATCAA 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 GGACGATGCCAGGCCAAGGTGTCACATCACCATGGTGGCCCTCAA 73
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-10-027-632-151846/c
; Sequence 151846, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
```





```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 789
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000215.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.2
; US-09-864-761-789

Query Match          10.6%; Score 31.8; DB 9; Length 468;
Best Local Similarity 51.8%; Pred. No. 1.1;
Matches    72; Conservative      0; Mismatches   67; Indels     0; Gaps     0;

QY      112 AGCAGGAGGCATGCGCAAGCTGGATAACTTGACCTTGTATGGATCAACACCAGCGGA 171
        ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       106 AGCAGGCGCAAGGACACCTTCGAGAAGGAGCTGCGTATCAGCACCTCACCGTGGAGG 165

QY      172 CTTTTCCTCACAAAGCGGCTCAACCACTATGACAACTCCGACAGAACCTCCAGCCTCTGG 231
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       166 CCTTGTGCTCAGGAGCGGAGCGACGCTCTCTGAGCTCCAGAAGGACCTTGGCGCTGAGA 225

QY      232 AAAGACCTCAGTAGACT 250
        | | | | | | | | | |
Db       226 AGAGCGGCACCTGGAGCT 244

Search completed: October 14, 2003, 17:17:50
Job time : 149.333 secs

```

Search completed: October 14, 2003, 17:17:50  
Job time : 149.333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 13:05:16 ; Search time 36.6667 Seconds  
(without alignments)  
3611.317 Million cell updates/sec

Title: US-09-854-124-7  
Perfect score: 300  
Sequence: 1 atcatgcttcagacaacatc.....tggggattaaccttcagac 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	11.7	2241	2	US-08-838-219B-20
2	35	11.7	2241	3	US-09-233-336A-20
3	35	11.7	2241	3	US-09-233-752A-20
4	35	11.7	2241	3	US-09-402-036-20
5	35	11.7	2241	4	US-09-904-226-20
6	35	11.7	2370	2	US-08-838-219B-19
7	35	11.7	2370	3	US-09-233-336A-19
8	35	11.7	2370	3	US-09-233-752A-19
9	35	11.7	2370	3	US-09-402-036-19
10	35	11.7	2370	4	US-09-904-226-19
11	35	11.7	2403	1	US-08-471-033-30
12	35	11.7	2403	2	US-08-471-044-30
13	35	11.7	2403	2	US-08-463-483A-30
14	35	11.7	2403	2	US-08-471-046A-30
15	35	11.7	2403	2	US-08-470-568B-30
16	35	11.7	2403	2	US-08-838-219B-7
17	35	11.7	2403	2	US-08-469-334-30
18	35	11.7	2403	3	US-09-300-529-30
19	35	11.7	2403	3	US-09-233-336A-7
20	35	11.7	2403	3	US-09-233-752A-7
21	35	11.7	2403	3	US-09-402-036-7
22	35	11.7	2403	4	US-09-904-226-7
23	32.6	10.9	1296	4	US-09-252-991A-7149
24	32.6	10.9	1419	4	US-09-252-991A-7347
25	32.6	10.9	2748	4	US-09-252-991A-7208
26	31.6	10.5	408	4	US-09-252-991A-9970
27	31.6	10.5	900	4	US-09-252-991A-10183

28 31.6 10.5 1332 4 US-09-252-991A-9889 Sequence 9889, Ap  
29 31 10.3 717 4 US-09-513-783A-43 Sequence 43, Appl  
30 31 10.3 1623 4 US-09-513-783A-33 Sequence 33, Appl  
31 31 10.3 2439 4 US-09-513-783A-3 Sequence 3, Appl  
32 31 10.3 2439 4 US-09-513-783A-5 Sequence 5, Appl  
33 31 10.3 3171 4 US-09-513-783A-31 Sequence 31, Appl  
34 31 10.3 4833 4 US-09-513-783A-21 Sequence 21, Appl  
35 30.8 10.3 1625 2 US-08-709-923-2 Patent No. 5472691  
36 29.8 9.9 666 6 5472691-4 Patent No. 5472691  
37 29.8 9.9 669 3 US-08-556-965-1 Patent No. 5472691  
38 29.8 9.9 723 6 5472691-7 Patent No. 5472691  
39 29.8 9.9 1389 2 US-08-023-980B-3 Sequence 3, Appl  
40 29.8 9.9 1389 2 US-08-486-953A-3 Sequence 3, Appl  
41 29.8 9.9 1396 6 5472691-1 Patent No. 5472691  
42 29.8 9.9 10079 2 US-08-476-866-20 Sequence 20, Appl  
43 29.6 9.9 4403765 3 US-09-103-840A-2 Sequence 2, Appl  
44 29.6 9.9 4411529 3 US-09-103-840A-1 Sequence 1, Appl  
45 29.4 9.8 717 4 US-09-513-783A-49 Sequence 49, Appl

## ALIGNMENTS

## RESULT 1

US-08-838-219B-20

; Sequence 20, Application US/08838219B  
; Patent No. 5877012

; GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W

; APPLICANT: Kozziel, Michael G

; APPLICANT: Mullins, Martha A

; APPLICANT: Nye, Gordon J

; APPLICANT: Carr, Brian

; APPLICANT: Desai, Nalini M

; APPLICANT: Kostichka, N. Kristy

; APPLICANT: Duck, Nicholas B

; APPLICANT: Estruch, Juan J

; TITLE OF INVENTION: A No. 5877012el Class of Proteins for the

; TITLE OF INVENTION: Control of Plant Pests

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/838,219B

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/463,483

; FILING DATE: 06-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/314,594

; FILING DATE: 09-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/037,057

; FILING DATE: 25-MAR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Pace, Gary M.

; REGISTRATION NUMBER: 40,403

; REFERENCE/DOCKET NUMBER: CGC 1925

; TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2241 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA encoding
; DESCRIPTION: VIP3A(c)"
; HYPOTHETICAL: NO
;
US-08-838-219B-20

Query Match 11.7%; Score 35; DB 2; Length 2241;
Best Local Similarity 59.6%; Pred. No. 0.14;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 107 GAGACAGCAGGAGCAGATGCCAAGCTGGGATAACTTGACCTTGATGGAGATCAACACGAG 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1950 GAGCCAGAACGGCGACGAGGCGCTGGGCGACAACCTTCATCATCTCTGGAGATCAGCCGAG 2009

QY 167 CGGACTTTCTCTCACACAAGCGCTCAACACCATGTACAA 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2010 CGAGAAGCTGCTGAGCCCGGAGCTGATCAACACCAACAA 2048

RESULT 2
US-09-233-336A-20
; Sequence 20, Application US/09233336A
; Patent No. 6107279
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: A No. 6107279el Class of Proteins for the
; TITLE OF INVENTION: Control of Plant Pests
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/233,336A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/838,219
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:

;
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1925
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2241 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA encoding
; DESCRIPTION: VIP3A(c)"
; HYPOTHETICAL: NO
;
US-09-233-336A-20

Query Match 11.7%; Score 35; DB 3; Length 2241;
Best Local Similarity 59.6%; Pred. No. 0.14;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 107 GAGACAGCAGGAGCAGATGCCAAGCTGGGATAACTTGACCTTGATGGAGATCAACACGAG 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1950 GAGCCAGAACGGCGACGAGGCGCTGGGCGACAACCTTCATCATCTCTGGAGATCAGCCGAG 2009

QY 167 CGGACTTTCTCTCACACAAGCGCTCAACACCATGTACAA 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2010 CGAGAAGCTGCTGAGCCCGGAGCTGATCAACACCAACAA 2048

RESULT 3
US-09-233-752A-20
; Sequence 20, Application US/09233752A
; Patent No. 6137033
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: A No. 6137033el Class of Proteins for the
; TITLE OF INVENTION: Control of Plant Pests
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/233,752A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/838,219
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
```

```

; GENERAL INFORMATION:
; APPLICANT: Estruch, Juan J.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Desai, Nalini
; APPLICANT: Kozziel, Michael
; APPLICANT: Nye, Gordon
; TITLE OF INVENTION: Plant Pest Control
; FILE REFERENCE: S-21284D
; CURRENT APPLICATION NUMBER: US/09/904,226
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/EP98/01952
; PRIOR FILING DATE: 1998-04-02

```

```

; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,263
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,265
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/463,483
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/314,594
; PRIOR FILING DATE: 1994-09-09
; PRIOR APPLICATION NUMBER: 08/218,018
; PRIOR FILING DATE: 1994-03-24
; PRIOR APPLICATION NUMBER: 08/037,057
; PRIOR FILING DATE: 1993-03-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; OTHER INFORMATION: encoding VIP3A(c)
US-09-904-226-20

Query Match      11.7% Score 35; DB 4; Length 2241;
Best Local Similarity 59.6%; Pred.No. 0.14;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY    107   GAGCAGCAGGAGGCACATGCCAACTGGGATAACTTGACCTTGTATGAGATCAACCAG 166
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     1950   GAGCCAGAGCGGACGAGGCGCTGGGCGCAAACCTTCATCATCTCTGGAGATCAGCCGAG 2009
               ||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY     167   CGGGACTTTCTCACACAAGCGCTCAACCAACATGTACAA 205
               ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db     2010   CGAGAAGCTCTGAGCCCGGAGCTGATCAACCAACA 2048

```

; Sequence 19, Application US/08838219B  
; Patent No. 5877012

## ; GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: A No. 5877012el Class of Proteins for the  
; TITLE OF INVENTION: Control of Plant Pests  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/838,219B  
; FILING DATE:  
; CLASSIFICATION: 530

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/463,483  
; FILING DATE: 06-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:

; NAME: Pace, Gary M.  
; REGISTRATION NUMBER: 40,403  
; REFERENCE/DOCKET NUMBER: CGC 1925  
; TELEPHONE: 919-541-8582  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 19:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 2370 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "synthetic DNA encoding  
; DESCRIPTION: VIP3A(b)"  
; HYPOTHETICAL: NO  
US-08-838-219B-19

Query Match 11.7%; Score 35; DB 2; Length 2370;  
Best Local Similarity 59.6%; Pred. No. 0.14;  
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 107 GAGACACGAGGCGACATCGCAGCTGGGATACCTTGCACCTTGATGGAGATCAACACGAG 166

Db 1950 GAGCCAGACGGCGACGAGGCGCTGGGCGGACAACTTCATCATCTCGGAGATCAGCCGAG 2009

QY 167 CGGAGCTTCTCTCACAAAGCGCTCAACACCATGTACAA 205

Db 2010 CGAGAAGCTGTGAGCCCGGAGCTGTGATCAACACCAACAA 2048

## RESULT 7

US-09-233-336A-19  
; Sequence 19, Application US/09233336A  
; Patent No. 6107279  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: A No. 6107279el Class of Proteins for the  
; TITLE OF INVENTION: Control of Plant Pests  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/233,336A  
; FILING DATE:  
; CLASSIFICATION:

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/838,219  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:

; NAME: Pace, Gary M.  
; REGISTRATION NUMBER: 40,403  
; REFERENCE/DOCKET NUMBER: CGC 1925  
; TELEPHONE: 919-541-8582  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 19:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 2370 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "synthetic DNA encoding  
; DESCRIPTION: VIP3A(b)"  
; HYPOTHETICAL: NO  
US-09-233-336A-19

Query Match 11.7%; Score 35; DB 3; Length 2370;  
Best Local Similarity 59.6%; Pred. No. 0.14;  
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 107 GAGACACGAGGCGACATCGCAGCTGGGATACCTTGCACCTTGATGGAGATCAACACGAG 166

Db 1950 GAGCCAGACGGCGACGAGGCGCTGGGCGGACAACTTCATCATCTCGGAGATCAGCCGAG 2009

US-09-904-226-19  
; Sequence 19, Application US/09904226  
; Patent No. 6429360  
; GENERAL INFORMATION:  
; APPLICANT: Estruch, Juan J.  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Desai, Nalini

```
; APPLICANT: Koziel, Michael
; APPLICANT: Nye, Gordon
; TITLE OF INVENTION: Plant Pest Control
; FILE REFERENCE: S-21284D
; CURRENT APPLICATION NUMBER: US/09/904,226
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/EP98/01952
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 08/838,219
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,263
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,265
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/463,483
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/314,594
; PRIOR FILING DATE: 1994-09-09
; PRIOR APPLICATION NUMBER: 08/218,018
; PRIOR FILING DATE: 1994-03-24
; PRIOR APPLICATION NUMBER: 08/037,057
; PRIOR FILING DATE: 1993-03-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; OTHER INFORMATION: encoding VIP3A(b)
; US-09-904-226-19

Query Match      11.7%; Score 35; DB 4; Length 2370;
Best Local Similarity 59.6%; Pred. No. 0.14;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 107 GAGACAGCAGGCGACATGCCAAGCTGGGATGAACCTTGACCTTGATGGAGATCAACACCG 166
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1950 GAGCGAGAACGGCGACGAGGCGCTGGGCGGACAACTTCATCATCTCGGAGATCAGCCGCG 2009

Qy 167 CGGACATTTCTCACAACGAGGCTCAACACCATGTACAA 205
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2010 CGAGAGCTGCTGAGCGCGGAGCTGATCAACACCAACAA 2048

RESULT 11
; Sequence 30, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11..2389
; OTHER INFORMATION: /note= "maize optimized DNA"
; OTHER INFORMATION: sequence encoding VIP3A(a)"
; US-08-471-033-30

Query Match      11.7%; Score 35; DB 1; Length 2403;
Best Local Similarity 59.6%; Pred. No. 0.14;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 107 GAGACAGCAGGCGACATGCCAAGCTGGGATGAACCTTGACCTTGATGGAGATCAACACCG 166
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1966 GAGCGAGAACGGCGACGAGGCGCTGGGCGGACAACTTCATCATCTCGGAGATCAGCCGCG 2025

Qy 167 CGGACATTTCTCACAACGAGGCTCAACACCATGTACAA 205
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2026 CGAGAGCTGCTGAGCGCGGAGCTGATCAACACCAACAA 2064

RESULT 12
US-08-471-044-30
; Sequence 30, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLv3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2403 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11..2389  
OTHER INFORMATION: /note= "maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIP3A(a)"  
US-08-471-044-30

Query Match 11.7%; Score 35; DB 2; Length 2403;  
Best Local Similarity 59.6%; Pred. No. 0.14;  
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 107 GAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCAG 166  
Db 1966 GAGCCAGAACGGCGAGGAGGCTGGGGCGACAACTTCATCTCTGGAGATCAGCCCGAG 2025  
QY 167 CGGGACTTTCCTCACACAAAGCGCTCAACACATGTACAA 205  
Db 2026 CGAGAAGCTGCTGAGCCCGAGAGCTGATCAACACCACAA 2064

RESULT 13  
US-08-463-483A-30  
Sequence 30, Application US/08463483A  
Patent No. 5849870  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Kozziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,483A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Spruill, W. Murray  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8615  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2403 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11..2389  
OTHER INFORMATION: /note= "maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIP3A(a)"  
US-08-463-483A-30

Query Match 11.7%; Score 35; DB 2; Length 2403;  
Best Local Similarity 59.6%; Pred. No. 0.14;  
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 107 GAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGACCTTGATGGAGATCAACACCAG 166  
Db 1966 GAGCCAGAACGGCGAGGAGGCTGGGGCGACAACTTCATCTCTGGAGATCAGCCCGAG 2025  
QY 167 CGGGACTTTCCTCACACAAAGCGCTCAACACATGTACAA 205  
Db 2026 CGAGAAGCTGCTGAGCCCGAGAGCTGATCAACACCACAA 2064

RESULT 14  
US-08-471-046A-30  
Sequence 30, Application US/08471046A  
Patent No. 5866326  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Kozziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy



Db 2026 CGAGNAGCTGCTGAGCCCGGAGCTGATCAACACCAACAA 2064

Search completed: October 14, 2003, 17:04:30  
Job time : 38.6667 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 12:04:45 ; Search time 157.333 Seconds  
(without alignments)  
5147.234 Million cell updates/sec

Title: US-09-854-124-7  
Perfect score: 300  
Sequence: 1 atcatgcttcagacaacatc.....tggggattaaccccttcagac 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	300	100.0	300	24	Differentially exp
2	244.4	81.5	858	22	Human cDNA sequenc
3	244.4	81.5	1171	25	Human NF-kappaB as
4	244.4	81.5	1210	21	Human cancer assoc
5	243	81.0	667	22	Human cDNA clone (
6	214.4	71.5	300	24	Differentially exp
7	205.4	68.5	678	23	DNA encoding novel
8	205.4	68.5	678	23	DNA encoding novel

9	194	64.7	1190	23	AAS871106	DNA encoding novel	
10	194	64.7	1499	23	AAS871104	DNA encoding novel	
11	158	52.7	454	24	ABL68902	Kidney cancer rela	
12	130.6	43.5	603	24	ABO59992	Human colon cancer	
13	75.2	25.1	649	24	ABO59972	Human colon cancer	
14	59.6	19.9	612	23	ABL19607	Drosophila melanog	
15	59.6	19.9	2670	23	ABL19606	Drosophila melanog	
16	59.6	19.9	4904	23	ABL05636	Drosophila melanog	
c	17	44.6	14.9	51	22	Human SNP oligonuc	
18	42	14.0	1257	23	AAS83051	DNA encoding novel	
19	40	13.3	940	23	AAS72563	DNA encoding novel	
20	40	13.3	1143	25	ABX71197	Novel human cDNA s	
21	39.6	13.2	734	22	AAH08244	Human cDNA clone (	
c	22	39.6	13.2	3071	22	AAH16628	Human cDNA sequenc
23	39	13.0	50	22	AAH16628	Human SNP oligonuc	
24	36	12.0	654	24	ABN61061	Human cancer relat	
25	35.8	11.9	1935	22	AAF60959	P. putida KT2440-a	
26	35	11.7	2241	19	AAV68070	Maize-optimised DN	
27	35	11.7	2370	19	AAV68069	Maize-optimised DN	
28	35	11.7	2403	17	AAT13955	Maize-optimised VI	
29	35	11.7	2403	18	AAAT74004	Maize-optimised-B.	
30	35	11.7	2403	19	AAV68065	Maize-optimised DN	
31	35	11.7	2403	19	AAV16177	Maize optimised DN	
32	35	11.7	2444	17	AAT13964	VIP3A(a) synthetic	
33	35	11.7	2444	18	AAT74011	B. cereus VIP3A(a)	
34	35	11.7	2444	19	AAV16190	DNA for maize opti	
35	33.8	11.3	3157	23	ABL13761	Drosophila melanog	
36	33.8	11.3	17330	23	ABL13760	Drosophila melanog	
37	33.4	11.1	5903	25	ABZ23939	Nucleotide sequenc	
c	38	32.6	10.9	39061	24	ABS78981	E. coli CFT073 gen
c	39	32.6	10.9	39061	24	ABS78985	E. coli CFT073 gen
40	32.4	10.8	5892	22	AAS00022	Human cDNA encodin	
41	32.4	10.8	6754	25	ABZ34760	Coding sequence SE	
42	32.4	10.8	7080	22	ABA08953	Human plexin-B1/SE	
43	32.2	10.7	2546	21	AAA09155	Human BMDSP-2 codi	
44	31.8	10.6	468	22	ABA42112	Human breast cell	
45	31.8	10.6	468	22	ABA52534	Human foetal liver	

ALIGNMENTS

RESULT 1

ABQ78752

XX

XX

AC

ABQ78752 standard; cDNA; 300 BP.

XX

ABQ78752;

XX

05-DEC-2002

(first entry)

XX

Differentially expressed gene RTA00000596F.d.12.1.

XX

Differentially expressed gene; cancer; breast cancer; colon cancer;

KW

lung cancer; prostate cancer; expressed sequence tag; EST;

KW

heat shock factor binding protein; CGI-122 gene; quiescin Q6;

KW

moderately-differentiated endometrial adenocarcinoma; TGIF protein;

KW

treacher collins syndrome protein; annexin IV; cyclophilin C;

KW

MHC class I lymphocyte antigen; HLA-E class I MRNA;

KW

glomeruloclerosis gene; 265 proteasome subunit p55;

KW

gamma interferon inducible protein; cyclin-dependent protein kinase;

KW

mitochondrial codecnyl-CoA delta-isomerase;

KW

serine hydroxymethyltransferase;

KW

DNA-damage-inducible RNA binding protein; ss.

XX

Homo sapiens.

OS

US2002076735-A1.

PN

20-JUN-2002.

PD

10-MAY-2001; 2001US-0854124.

PF

25-SEP-1998; 98US-101900P.

XX

PR

ALIGNMENTS

RESULT 1	ABQ78752	standard; cDNA; 300 BP.
ID	ABQ78752	standard; cDNA; 300 BP.
XX	AC	ABQ78752;
XX	DT	05-DEC-2002 (first entry)
XX	DE	Differentially expressed gene RTA00000596F.d.12.1.
XX	KW	Differentially expressed gene; cancer; breast cancer; colon cancer;
XX	KW	lung cancer; prostate cancer; expressed sequence tag; EST;
XX	KW	heat shock factor binding protein; CGI-122 gene; quiescin Q6;
XX	KW	moderately-differentiated endometrial adenocarcinoma; TGIF protein;
XX	KW	treacher collins syndrome protein; annexin IV; cyclophilin C;
XX	KW	MHC class I lymphocyte antigen; HLA-E class I MRNA;
XX	KW	glomerulosclerosis gene; 265 proteasome subunit p55;
XX	KW	gamma interferon inducible protein; cyclin-dependent protein kinase;
XX	KW	mitochondrial dodecenoyl-coA delta-isomerase;
XX	KW	serine hydroxymethyltransferase;
XX	KW	DNA-damage-inducible RNA binding protein; ss.
OS	Homo sapiens.	
XX	XX	US2002076735-A1.
PN	20-JUN-2002.	
XX	10-MAY-2001; 2001US-0854124.	
XX	25-SEP-1998; 98US-101900P.	

```

PR 22-SEP-1999; 99US-0400947.
XX (WILL/) WILLIAMS L T.
PA (ESCO/) ESCOBEDO J.
PA (INNI/) INNIS M A.
PA (GARC/) GARCIA P D.
PA (SUDD/) SUDDUTH-KLINGER J.
PA (REIN/) REINHARD C.
PA (GIES/) GIESE K.
PA (RAND/) RANDAZZO F.
PA (KENN/) KENNEDY G C.
PA (POTD/) POT D.
PA (KASS/) KASSAM A.
PA (LWAM/) LAMSON G.
PA (DRMA/) DRMANAC R.
PA (CRKV/) CRKVENJAKOV R.
PA (DICK/) DICKSON M.
PA (DRMA/) DRMANAC S.
PA (LABA/) LABAT I.
PA (LESH/) LESHKOWITZ D.
PA (KITA/) KITA D.
PA (GARC/) GARCIA V.
PA (JONE/) JONES L W.
PA (STAC/) STACHE-CRAIN B.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX WPI; 2002-673346/72.
XX
XX Detecting cancerous breast, colon, lung and prostate cells and
PT assessing metastatic potential of cells of the breast, colon and lung,
PT and preventing metastasis in cancer cells using nucleic acids
XX
XX Claim 1; Page 29; 38pp; English.
XX
XX ABQ78746-82 represents genes that are differentially expressed in
CC cancerous cells. Detection of the expression level of these genes, and
CC then comparing with the level of expression in a control cell, can be
CC used for detecting a cancerous breast, colon, lung or prostate cell, and
CC for assessing the metastatic potential of that cell. ABQ78746-47, and
CC ABQ78753-54 and ABQ78756 correspond to an expressed sequence tag (EST);
CC ABQ78748-49 correspond to a heat shock factor binding protein;
CC ABQ78750-52 correspond to CGI-122 gene; ABQ78755 corresponds to a CDNA
CC sequence isolated from moderately-differentiated endometrial
CC adenocarcinoma; ABQ78757 corresponds to quiescin Q6; ABQ78758
CC corresponds to treacher collins syndrome protein; ABQ78759 corresponds
CC to annexin IV; ABQ78760 corresponds to TgIF protein; ABQ78761
CC corresponds to MHC class I lymphocyte antigen; ABQ78762 corresponds to
CC HLA-E class I mRNA; ABQ78763 corresponds to glomerulosclerosis gene;
CC ABQ78764-66 correspond to cyclophilin C; ABQ78767 corresponds to 26S
CC proteasome subunit p55; ABQ78768-72 corresponds to gamma interferon
CC inducible protein; ABQ78773 corresponds to Na+/H-exchange regulatory
CC co-factor; ABQ78774-77 corresponds to mitochondrial dodecenoyl-CoA
CC delta-isomerase; ABQ78778 corresponds to cyclin-dependent protein
CC kinase; ABQ78779-81 correspond to serine hydroxymethyltransferase; and
CC ABQ78782 corresponds to DNA-damage-inducible RNA binding protein. The
CC method is used to detect monitor and treat cancer of the breast, colon,
CC lung and prostate.
XX
XX Query Match 100.0%; Score 300; DB 24; Length 300;
SQ Best Local Similarity 100.0%; Pred. No. 1e-86;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCATGCTTCAGACAACATCCGGAAGCAGACGAAATCCGGACCTGGTCAAGGATATGT 60
DB 1 ATCATGCTTCAGACAACATCCGGAAGCAGACGAAATCCGGACCTGGTCAAGGATATGT 60
QY 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 120

```

```

DB 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGG 120
QY 121 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACGCGGGACTTTCCTCA 180
DB 121 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACGCGGGACTTTCCTCA 180
QY 181 CACAAGCGCTCAACACACATGTACAAACTCCCGACGAACTCCAGCCTCGAAGAGACCTC 240
DB 181 CACAAGCGCTCAACACACATGTACAAACTCCCGACGAACTCCAGCCTCGAAGAGACCTC 240
QY 241 AGCTAGGACTTCTAAAAAGCGCTGGTGCAGCCGCTTGGTTGGGGATTAACCCCTTCAGAC 300
DB 241 AGCTAGGACTTCTAAAAAGCGCTGGTGCAGCCGCTTGGTTGGGGATTAACCCCTTCAGAC 300

RESULT 2
AAH14111
ID AAH14111 standard; cDNA; 858 BP.
XX
XX AAH14111;
AC
XX
XX 26-JUN-2001 (first entry)
DT
XX
XX Human cDNA sequence SEQ ID NO:11295.
DE
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
FA
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 11295; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC

```



08-FEB-2001 (first entry)  
Human cancer associated gene sequence SEQ ID NO:214.  
Human: cancer associated gene; cancer antigen; detection; cancer; diagnosis; cytostatic; proliferative; vulnarary; immunomodulator; antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral; antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiant; dermatologic; neuroprotective; thrombolytic; coagulant; nootropic; vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation; immune disorder; haematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus host disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurological disease; drug screening; ss.  
Homo sapiens.  
WO200053350-A1.  
21-SEP-2000.  
08-MAR-2000; 2000WO-US05882.  
12-MAR-1999; 99US-0124270.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Ruben SM;  
WPI: 2000-587533/55.  
P-FSDB; AAB43611.  
Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer -  
Claim 1; Page 785-786; 2352pp; English.  
AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulnarary; immunomodulator; antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiant; dermatologic; neuroprotective; thrombolytic; coagulant; nootropic; vasotropic; antipsoriatic and antiangiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention.  
Sequence 1210 BP; 320 A; 300 C; 320 G; 268 T; 2 other;

Query Match 81.5%; Score 244.4; DB 21; Length 1210;  
Best Local Similarity 95.1%; Pred. No. 1.6e-68;  
Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACAACATCCGAGGACGACGAAATCCGACCTGGTCAAGGATATGT 60  
DB 382 ATTCATGCTTCAGACAACATCCGAGGACGACGAAATCCGACCTGGTCAAGGATATGT 441  
QY 61 GGGACACTCTATAGCCAAACTCCGAGTGTCTGTGACAGTGTGTGAGACAGCAGGAGG 120  
DB 442 GGGACACTCTATAGCCAAACTCCGAGTGTCTGTGACAGTGTGTGAGACAGCAGGAGG 501

QY 121 CACATGCCAAGCTGGATAAATCTGACCTTGATGGAGATCAACACACGCGGACCTTCTCTCA 180  
DB 502 CACATGCCAAGCTGGATAAATCTGACCTTGATGGAGATCAACACACGCGGACCTTCTCTCA 561  
QY 181 CACAAGCGCTCAACCCACATGTACAACTCCGACGAACTCCAGCCTCTCGAAGACCTC 240  
DB 562 CACAAGCGCTCAACCCACATGTACAACTCCGACGAACTCCAGCCTCTCGAAGACCTC 621  
QY 241 AG--CTAGGACTTCTTA-AAAAAGCCCTGGTGCAGCCCTGGTTGGGG 285  
DB 622 AGTCTCAGGACTTCTTAGAAGAAAGCCCTGGTGCAGCCCTGGTTGGGG 669

RESULT 5  
AAH05624  
ID AAH05624 standard; cDNA; 667 BP.  
AC AAH05624;  
XX 26-JUN-2001 (first entry)  
XX Human cDNA clone (5'-primer) SEQ ID NO:2459.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318745/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 1; SEQ ID 2459; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesising 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesising polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632



Qy 268 GCAGCCGCTTGCTT

```

Db      241 GCAGGGCGCTGCTGGGG 258
      |||| | | | | | | | |
RESULT 7
AAS87103
ID AAS87103 standard; cDNA; 678 BP.
XX
AC AAS87103;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #22907.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG22916.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID No 22907; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 678 BP; 177 A; 185 C; 206 G; 110 T; 0 other;
XX
Query Match 68.58; Score 205.4; DB 23; Length 678;
Best Local Similarity 91.08; Pred. No. 5e-56;
Matches 252; Conservative 0; Mismatches 21; Indels 4; Gaps 3;
Oy 12 GACAACTCCGAGGAGGAGCAATCCGGACCTGCTCAAGGATGTGGGACACTCGT 71
Db 382 GACAACTCCGAGGAGGAGCAATCCGGACCTGCTCAAGGATGTGGGACACTCGT 441

```

```

Oy 72 ATAGCCAAACTCCGAGTCTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAG 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 442 ATAGCCAAACTCCGAGTCTGCTGACAGCTTTGTGAGGCAACAGAGGCACATGCCAAG 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 132 CTGGATAACTTGACCTTGATGGAGATCAACACCAGCGGGACTTTCTCTCACACAAGCGCTC 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 502 CCGGATAACTTGACCTTGATGGAGATCAACCCAGCGGGACTTTCTCTCACACAAGCGCTC 561
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 192 AACCATGTGTACAACTCCGACGAACTCCAGCCTCTGGAAGACCTCAG--CTAGGAC 249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 562 AACCATGTGTACAACT--CGCACAACTCCAGCCTCTCGGAGAGTACTCAGTCTCAGGAC 620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 250 TTCTA-AAAAAGCGCTGTCAGCGCTTGTTGGGG 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 621 TTCTAGAGAAAGCGCTGTCAGCGCTTGTTGGGG 657
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 8
AAS87105
ID AAS87105 standard; cDNA; 678 BP.
XX
AC AAS87105;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #22909.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG22916.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID No 22909; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC amino acid sequences. AAS64197-AAS94564 represent novel human

```

CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 678 BP; 177 A; 185 C; 206 G; 110 T; 0 other;  
 Query Match 68.5%; Score 205.4; DB 23; Length 678;  
 Best Local Similarity 91.0%; Pred. No. 5e-56;  
 Matches 252; Conservative 0; Mismatches 21; Indels 4; Gaps 3;  
 QY 12 GACACATCCCGAAGGCACACGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCCT 71  
 DB 382 GACAACATCCCGAAGGCACACGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCCT 441  
 QY 72 ATAGCCAAACTCCGAGTGTCTGTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAG 131  
 DB 442 ATAGCCAAACTCCGAGTGTCTGTGACAGCTTTGTGAGGCAACAGAGGCACATGCCAAG 501  
 QY 132 CTGGATACTTACCTTGATGGAGATCAACACAGCGGGACTTTCCTCACAAAGCGCTC 191  
 DB 502 CCGGATACTTACCTTGATGGAGATCAACACAGCGGGACTTTCCTCACAAAGCGCTC 561  
 QY 192 AACCATCTACAACTCCGACGACCTCCAGCCTCTGGAAGACCTCAG--CTAGGAC 249  
 DB 562 AACCATCTACAACT--CGCACAACTCCAGCCTCTCGGAGAGTACTCAGTCTCAGGAC 620  
 QY 250 TTCTA-AAAAAGCCTGTGTCAGCCGCTTTGGTTGGGG 285  
 DB 621 TTCTAGAAAGCCTGTGTCAGCCAGCTTGTCTGGGG 657  
 RESULT 9  
 AAS87106  
 ID AAS87106 standard; cDNA; 1190 BP.  
 AC AAS87106;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #22910.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 PD 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR P-PSDB; ABG22919.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 1; SEQ ID No 22910; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1190 BP; 298 A; 304 C; 310 G; 278 T; 0 other;  
 Query Match 64.7%; Score 194; DB 23; Length 1190;  
 Best Local Similarity 92.5%; Pred. No. 3e-52;  
 Matches 270; Conservative 0; Mismatches 15; Indels 7; Gaps 6;  
 QY 1 ATCATGCTTCAGACAAACATCCCGAAGGCAGACGAAATCCGGACCCCTGGTCAAGGATATGT 60  
 DB 383 ATCATGCTTCAGACAAACATCCCGAAGGCAGACGAAATCCGGACCCCTGGTCAAGGATATGT 442  
 QY 61 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACAGC--TTTGTGAGACAGCAGGAG 119  
 DB 443 GGGACACTCGTATAGCCAAACTCCGAGTGTCTGTGACAGCTTTTGTGAGACAGCAGGAG 502  
 QY 120 GCACATGCCAAGCT--GGATAACTTGACCTT--GATGGAGATCAACACCGCGG--GACTTTC 176  
 DB 503 GCACATGCCAAGCTCGGATAACTTGACTTTCGATGGAGATCCACACCGCGGTGACTTTC 562  
 QY 177 CTCACAAAGCGCTCAACCCACATGTACAACTCCGCACGAACTCCAGCCTCTCGGAAAGA 236  
 DB 563 CTCACAAAGCGCTCAACCCACATGTACAACTCCGCACGAACTCCAGCCTCTCGGAGAGT 622  
 QY 237 CCTCAG--CTAGGACTTCTA-AAAAAGCCTGTGTCAGCGCTTGTGGGG 285  
 DB 623 ACTCAGTCTCAGGACTTCTAGAGAAAGGCTGTGTCAGCGCGCTTGTGGGG 674  
 RESULT 10  
 AAS87104  
 ID AAS87104 standard; cDNA; 1499 BP.  
 AC AAS87104;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #22908.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 PD 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;



CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 454 BP; 128 A; 113 C; 116 G; 97 T; 0 other;

Query Match 52.7%; Score 158; DB 24; Length 454;  
Best Local Similarity 98.9%; Pred. No. 8.3e-41;  
Matches 180; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
QY 1 ATCATGCTTCACACACATCCCGAAGGCGAGACGAATCCGGACCCCTGGTCAAGGATATGT 60  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
274 ATCATGCTTCACACACATCCCGAA-GCAGACGAATCCGGACCCCTGGTCAAGGATATGT 332  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACAGCAGGAGG 120  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
333 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACAGCAGGAGG 392  
QY 121 CACATGCCAAGCTGGATACCTTGACCTTGATGGAGATCAACACCAGCGGACTTTCTCTCA 180  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
393 CACATG-CAAGCTGGATACCTTGACCTTGATGGAGATCAACACCAGCGGACTTTCTCTCA 451  
QY 181 CA 182  
DB ||  
452 CA 453

RESULT 12  
ABQ59992  
ID ABQ59992 standard; cDNA; 603 BP.  
AC ABQ59992;  
XX  
DT 02-AUG-2002 (first entry)  
XX  
DE Human colon cancer related nucleotide sequence SEQ ID NO:3687.  
XX  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200229086-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US30732.  
XX  
XX 02-OCT-2000; 2000US-237271P.  
XX  
XX (FARB ) BAYER CORP.  
XX  
XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thiagalingam A, Lewis ME;  
XX  
XX WPI; 2002-426115/45.  
XX  
XX New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell  
PT or tissue type, and in antisense therapy -  
XX  
XX Claim 1; Fig 1; 796pp; English.  
XX  
XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be

CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample, and  
CC for detecting the presence or absence of a polynucleotide encoded by a  
CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridises to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists.  
XX  
SQ Sequence 603 BP; 157 A; 161 C; 162 G; 117 T; 6 other;

Query Match 43.5%; Score 130.6; DB 24; Length 603;  
Best Local Similarity 90.6%; Pred. No. 6.5e-32;  
Matches 184; Conservative 0; Mismatches 11; Indels 8; Gaps 4;  
QY 1 ATCATGCTTCACACACATCCCGAAGGCGAGACGAATCCGGACCCCTGGTCAAGGATATGT 60  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
364 ATCATGCTTCACACACATCCCGAAGGCGAGACGAATCCGGACCCCTGGTCAAGGATATGT 423  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACAGCAGGAGG 120  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
424 GGGACACTCGTATAGCCAACTCCGA-CGTCTGCTGACAGCTTTGTGACAGCAGGAGG 482  
QY 121 CACATGCCAAGCTGGATACCTTGACCTTGATGGAGATCAACACCAGCGGACTTT 173  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
483 CACATGCCAAGCTGGATACCTTGACCTTGATGGAGATCAACACCAGCGGACTTT 542  
QY 174 TTCCTCACACAAAGCGCTCAACCA 196  
DB |||||||||||||||||||||||||||  
543 TCCTTCACACAAAGCGCTCAACCA 565

RESULT 13  
ABQ59972  
ID ABQ59972 standard; cDNA; 649 BP.  
XX  
AC ABQ59972;  
XX  
DT 02-AUG-2002 (first entry)  
XX  
DE Human colon cancer related nucleotide sequence SEQ ID NO:3667.  
XX  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200229086-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US30732.  
XX  
XX 02-OCT-2000; 2000US-237271P.  
XX  
XX (FARB ) BAYER CORP.  
XX  
XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thiagalingam A, Lewis ME;  
XX  
XX WPI; 2002-426115/45.  
XX  
XX New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell  
PT or tissue type, and in antisense therapy -  
XX  
XX Claim 1; Fig 1; 796pp; English.  
XX

CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
 CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins  
 CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
 CC encoded by (I) is useful for detecting cancer in a patient sample, and  
 CC for detecting the presence or absence of a polynucleotide encoded by a  
 CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
 CC from (I) can be used for determining the presence of a nucleic acid which  
 CC hybridises to (I), and for determining the phenotype of cells in a sample  
 CC of cells from a patient. (I) is useful for determining the presence of  
 CC colon cancer in a cell or tissue type, for determining the presence or  
 CC state of other type of cancer, in antisense therapy, to generate  
 CC macroarrays on a solid surface, to identify a chromosome on which the  
 CC corresponding gene resides, and in tissue profiling, forensics, genetic  
 CC analysis, mapping and diagnostic applications. (I) can be used to raise  
 CC antibodies, and to screen for peptide analogues and antagonists.

XX Sequence 649 BP; 159 A; 163 C; 169 G; 138 T; 20 other;

Query Match 25.1%; Score 75.2; DB 24; Length 649;

Best Local Similarity 76.1%; Pred. No. 5e-14;

Matches 150; Conservative 0; Mismatches 39; Indels 8; Gaps 5;

QY 9 TCAGACACATCCCGAAGGCAGACGAAATCCGGACCTTGTCAGGATATGTGGACACT 68

DB 374 TTNNACACATCCNGAAGGCANACNAAATCCGGACCTTGTCAGGATATGTGGACACT 433

QY 69 --CGTATAGCCAACTCCGAGTGTCTGC-TGACAGCTTTGTGAGAC-AGCAGGAGGCACA 124

DB 434 TCGTNTTAGCCAACTCCGAGTGTCTGCTTGACAGTTTTTNGANACAGCAGGAGNACA 493

QY 125 TGCCAAGCT--GGATACTTACCTT--GATGGAGATCAACACAGCGGGACTTTCCTCA 180

DB 494 TGCCAAGCTTGGATACTTGACCTTTGATGGANATCAACACCGCGGGACTTTTCTT 553

QY 181 CACAAGCCTCAACAC 197

DB 554 AANACAGCCCTTCAAC 570

# RESULT 14

ABL19607

ID ABL19607 standard; DNA; 612 BP.

AC ABL19607;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 10294.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 10294; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).

CC

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 612 BP; 155 A; 166 C; 153 G; 138 T; 0 other;

Query Match 19.9%; Score 59.6; DB 23; Length 612;

Best Local Similarity 56.7%; Pred. No. 5.3e-09;

Matches 110; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 6 GTTTCAGACACATCCCGAAGGCAGACGAAATCCGGACCTTGTCAGGATATGTGGAC 65

DB 307 GCTCCGATGACGTTCGCGTTGCGAGGAGCTGCGCACTGTGATCAAGGACATTTTCGAC 366

QY 66 ACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAGGCACAT 125

DB 367 ATACGCGAGTCCAAAGCTGCGCACTTCGATCGACGCCCTTTATCAAGGAGAGGGCACATAT 426

QY 126 GCCAAGCTGGATACCTTGACCTTGATGGAGATCAACACCAGCGGGACTTTCCTCACACAA 185

DB 427 GCAAGGTAGACAACTCAACCTTCTGGAGATCCACAGCGGTGAGACCCATTTCTGCCCTAT 486

QY 186 GCGCTCAACCCACAT 199

DB 487 TCCCTGGACCCACAT 500

# RESULT 15

ABL19606

ID ABL19606 standard; DNA; 2670 BP.

AC ABL19606;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 10291.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 10291; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from prosochila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2670 BP; 733 A; 527 C; 535 G; 675 T; 0 other;  
  
Query Match 19.9%; Score 59.6; DB 23; Length 2670;  
Best Local Similarity 56.7%; Pred. No. 9.8e-09;  
Matches 110; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
  
Qy 6 GCTTCAGACACATCCCGAGGCGACGAGCAATCCGGACCCCTGGTCAAGGATATGTGGAC 65  
Dy 1365 GCTCCGGATGACGTCCCGCGTTGCGAGGAGCTGGCGACCTGTGATCAAGGACATCTCGAC 1424  
Qy 66 ACTCGTATAGCCAAACTCCGAGTCTCTGCTGACAGCTTTGTGAGACAGCAGGAGGCACAT 125  
Dy 1425 ATACGCGAGTCCAAGCTGGCGACTTCGATCGACGCCCTTTATCAAGGAGAGGGGCACATAT 1484  
Qy 126 GCCAAGCTGGATACTTGACCTTGATGGAGATCAACACACCGGGGACTTTCTCCTCACACAA 185  
Dy 1485 GCAAAGCTAGACAACCTCACGCTTCTGGAGATCCACACGCGTGACACCCATTCGCCCTAT 1544  
Qy 186 GCGCTCAACACAT 199  
Dy 1545 TCCTGGACCACAT 1558

Search completed: October 14, 2003, 14:17:04  
Job time : 158.333 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 14, 2003, 13:00:56 ; Search time 1477 Seconds  
(without alignments)  
8309.335 Million cell updates/sec

Title: US-09-854-124-7  
Perfect score: 300  
Sequence: 1 atcatgcttcagacaacatc.....tggggattacccttcagac 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_man.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	244.4	81.5	858	6	BD156103	BD156103 Primer fo
2	244.4	81.5	858	9	AK001275	AK001275 Homo sapi
3	244.4	81.5	1174	9	AF151880	AF151880 Homo sapi
4	244.4	81.5	1180	9	BC010164	BC010164 Homo sapi
5	244.4	81.5	1188	9	AF201939	AF201939 Homo sapi
6	243	81.0	667	6	BD147616	BD147616 Primer fo
7	242.8	80.9	1203	9	BC003186	BC003186 Homo sapi
c 8	216.4	72.1	116792	9	HS1100H13	AL035419 Human DNA
9	158	52.7	454	6	AX336730	AX336730 Sequence
10	136.4	45.5	933	5	AB097169	AB097169 xenopus 1
c 11	128.2	42.7	160356	9	AC009068	AC009068 Homo sapi
c 12	128.2	42.7	185321	9	AC123908	AC123908 Homo sapi
13	128.2	42.7	245210	2	AC137771	AC137771 Homo sapi
14	126.6	42.2	3097	9	AK091519	AK091519 Homo sapi
c 15	99.6	33.2	188292	2	AC103360	AC103360 Mus muscu
c 16	94	31.3	207307	2	AC117085	AC117085 Rattus no
17	59.6	19.9	825	3	AY071234	AY071234 Drosophill
c 18	59.6	19.9	72722	3	AC004373	AC004373 Drosophill
c 19	59.6	19.9	132666	3	AC005270	AC005270 Drosophill
20	59.6	19.9	182897	3	AC092232	AC092232 Drosophill
c 21	59.6	19.9	218794	2	AC018064	AC018064 Drosophill
c 22	59.6	19.9	292393	3	AE003576	AE003576 Drosophill
23	47.2	15.7	213	9	AF534827	AF534827 Homo sapi
24	47.2	15.7	92814	9	AL391903	AL391903 Human DNA
25	47.2	15.7	219443	2	AC023320	AC023320 Homo sapi
26	44.6	14.9	165158	2	AC141687	AC141687 Apis mell
27	40	13.3	1150	9	AF534824	AF534824 Homo sapi
28	40	13.3	91511	9	AL133404	AL133404 Human DNA
c 29	39.6	13.2	734	6	BD150236	BD150236 Primer fo
30	39.6	13.2	3071	6	BD158620	BD158620 Homo sapi
31	39.6	13.2	3071	9	AK023755	AK023755 Homo sapi
32	36.2	12.1	225145	2	AC111766	AC111766 Rattus no
33	36.2	12.1	241188	2	AC133023	AC133023 Rattus no
c 34	36.2	12.0	654	6	AX386100	AX386100 Sequence
c 35	36	12.0	49995	2	AC011335	AC011335 Homo sapi
c 36	36	12.0	77139	2	AC008466	AC008466 Homo sapi
c 37	36	12.0	120849	2	AC087148	AC087148 Mus muscu
c 38	36	12.0	128928	9	AC068658	AC068658 Homo sapi
c 39	36	12.0	157208	9	AC113409	AC113409 Homo sapi
c 40	36	12.0	203418	2	AC063958	AC063958 Homo sapi
c 41	35.8	11.9	1935	6	AX078491	AX078491 Sequence
42	35.8	11.9	2746	8	NCACU9	X56672 N.crassa ma
43	35.6	11.9	1129	8	AF499724	AF499724 Theilungi
44	35.6	11.9	215707	2	AC099096	AC099096 Rattus no
c 45	35.6	11.9	215707	2	AC099096	AC099096 Rattus no

ALIGNMENTS

RESULT 1  
LOCUS BD156103 858 bp DNA linear PAT 17-JAN-2003  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION BD156103  
VERSION BD156103.1 GI:27861861  
KEYWORDS JP 2002191363-A/10946.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 858)  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuk,T.  
TITLE Primer for synthesizing full-length cDNA and use thereof

JOURNAL Patent: JP 2002191363-A 10946 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 COMMENT OS Homo sapiens (human)  
 PN JP 2002191363-A/10946  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OYA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N5/00, C12P21/02, C12P21/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 Location/Qualifiers  
 FT CDS (37)..(591).  
 FEATURES  
 source  
 1..858  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 227 a 214 c 222 g 195 t  
 ORIGIN  
 Query Match 81.5%; Score 244.4; DB 6; Length 858;  
 Best Local Similarity 95.1%; Pred. No. 9.6e-55;  
 Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;  
 QY 1 ATCATGCTTCACACACATCCCGAGGCGAGCAATCCGGACCTGGTCAAGGATATGT 60  
 DB 338 ATCATGCTTCACACACATCCCGAGGCGAGCAATCCGGACCTGGTCAAGGATATGT 397  
 QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 120  
 DB 398 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 457  
 QY 121 CACATGCCAAGCTGGATACTTGACCTTGATGAGATCAACACCAGCGGACTTTCTCTCA 180  
 DB 458 CACATGCCAAGCTGGATACTTGACCTTGATGAGATCAACACCAGCGGACTTTCTCTCA 517  
 QY 181 CACAAGCGCTCAACACATGTACAACTCCGACGAACTCCAGCCTCTGGAAGACCTC 240  
 DB 518 CACAAGCGCTCAACACATGTACAACTCCGACGAACTCCAGCCTCTGGAAGACCTC 577  
 QY 241 AG--CTAGGACTCTA-AAAAAGCGCTGTGTCAGCGCTTGGTTGGG 285  
 DB 578 AGTCTCAGGACTCTAGAGAAAGCGCTGTGTCAGCGGCTTGTCTGGG 625  
 RESULT 2  
 AK001275 858 bp mRNA linear PRI 01-AUG-2002  
 LOCUS Homo sapiens cDNA FLJ10413 fis, clone NT2RP1000063.  
 DEFINITION AK001275  
 ACCESSION AK001275.1 GI:7022428  
 VERSION AK001275.1 GI:7022428  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
 Nishikawa, T., Nagai, K., Sugeno, S., Shiratori, A., Sudo, H.,  
 Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
 Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,  
 Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,  
 Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,  
 Nakamura, Y., Nagahara, K., Masubo, Y., Ninomiya, K. and Iwayanagi, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 858)  
 Isogai, T. and Otsuki, T.  
 TITLE  
 JOURNAL  
 COMMENT  
 REFERENCE  
 AUTHORS  
 Direct Submission  
 Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection;  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 Location/Qualifiers  
 1..858  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP1000063"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP1"  
 /note="cloning vector: pUC19FL3-mRNA from NT2 neuronal  
 precursor cells after 48-hours retinoic acid (RA)  
 induction."  
 37..594  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAA91595.1"  
 /db\_xref="GI:7022429"  
 /translation="MDAARVEFLAEKELVTIIPNSLDKIYILGGDLGPFNPCLPVEV  
 PMLALNLRQKRCRLPPMDVKEKLRDHERKEETFTPMPSYIMELKLLNLH  
 ANDPKADIEITLVKMDWDTIAKLRVSADSFVRQOEAAKLDNLTLMEINTSGTFL  
 TQALNHYKLRITNLQPLESTQSDP"  
 BASE COUNT 227 a 214 c 222 g 195 t  
 ORIGIN  
 Query Match 81.5%; Score 244.4; DB 9; Length 858;  
 Best Local Similarity 95.1%; Pred. No. 9.6e-55;  
 Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;  
 QY 1 ATCATGCTTCACACACATCCCGAGGCGAGCAATCCGGACCTGGTCAAGGATATGT 60  
 DB 338 ATCATGCTTCACACACATCCCGAGGCGAGCAATCCGGACCTGGTCAAGGATATGT 397  
 QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 120  
 DB 398 GGGACACTCGTATAGCCAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGG 457  
 QY 121 CACATGCCAAGCTGGATACTTGACCTTGATGAGATCAACACCAGCGGACTTTCTCTCA 180  
 DB 458 CACATGCCAAGCTGGATACTTGACCTTGATGAGATCAACACCAGCGGACTTTCTCTCA 517  
 QY 181 CACAAGCGCTCAACACATGTACAACTCCGACGAACTCCAGCCTCTGGAAGACCTC 240  
 DB 518 CACAAGCGCTCAACACATGTACAACTCCGACGAACTCCAGCCTCTGGAAGACTC 577  
 QY 241 AG--CTAGGACTCTA-AAAAAGCGCTGTGTCAGCGCTTGGTTGGG 285  
 DB 578 AGTCTCAGGACTCTAGAGAAAGCGCTGTGTCAGCGGCTTGTCTGGG 625  
 RESULT 3  
 AF151880 1174 bp mRNA linear PRI 18-MAY-2000  
 LOCUS Homo sapiens CGI-122 protein mRNA, complete cds.  
 DEFINITION AF151880  
 ACCESSION AF151880  
 VERSION AF151880.1 GI:4929712  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1174)  
 Lai, C.H., Chou, C.Y., Ch'ang, L.Y., Liu, C.S. and Lin, W.

**TITLE** Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics

**JOURNAL** Genome Res. 10 (5), 703-713 (2000)

**MEDLINE** 20272150

**PUBMED** 10810093

**REFERENCE** 2 (bases 1 to 1174)

**AUTHORS** Lin, W.-C.

**TITLE** Direct Submission

**JOURNAL** Submitted (17-MAY-1999) Institute of Biomedical Sciences, Academia Sinica, No. 128, Sec. II, Academia Road, Taipei 115, Taiwan

**FEATURES** Location/Qualifiers

**source** 1..1174

**CDS** /organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="16"

77..634

/codon\_start=1

/product="CGI-122 protein"

/protein\_id="AAD34117.1"

/db\_xref="GI:4929713"

/translation="MDAAEEVEFLAEKELVTIIPNFSLDKIYILIGDLGPFNPGLPVEV PLWLAINKQKCKLLPEWMDVEKLEMRDHERKEETFTPMPSPYMELTKLLLNH ASDNPKADEIRTLVKDMDTRIAKLRVSADSFVQQAHAHAKLDNLTLMEINTSGTFL TQALNHYKLRNLPLESTQSQDF"

**BASE COUNT** 300 a 299 c 307 g 268 t

**ORIGIN**

Query Match 81.5%; Score 244.4; DB 9; Length 1174;

Best Local Similarity 95.1%; Pred. No. 9.2e-55;

Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACAACATCCGGAAGGCAGAGCAATCCGGACCTGGTCAAGGATATGT 60

|||||

DB 378 ATCATGCTTCAGACAACATCCGGAAGGCAGAGCAATCCGGACCTGGTCAAGGATATGT 437

QY 61 GGGACACTCGTATACCAACTCCGAGTGTCTGTGACAGCTTTGTGACAGCAGGAGG 120

|||||

DB 438 GGGACACTCGTATACCAACTCCGAGTGTCTGTGACAGCTTTGTGACAGCAGGAGG 497

QY 121 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACACCGCGGACTTTCCCTCA 180

|||||

DB 498 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACACCGCGGACTTTCCCTCA 557

QY 181 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTCGAAGACCTC 240

|||||

DB 558 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTCGAAGACTACTC 617

QY 241 AG--CTAGGACTTCTA-AAAAAGGCGCTGGTGCAGCGCTTGGTGGG 285

|||||

DB 618 AGTCTCAGGACTTCTAGAGAAAGGCGCTGGTGCAGCGCTTGGTGGG 665

**RESULT 4**

**BC010164** 1180 bp mRNA linear PRI 12-JUL-2001

**LOCUS** Homo sapiens, HSPC037 protein, clone MGC:19836 IMAGE:4098007, mRNA, complete cds.

**DEFINITION** BC010164

**ACCESSION** BC010164.1 GI:14603431

**VERSION** MGC.

**KEYWORDS** Homo sapiens (human)

**SOURCE** Homo sapiens

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 1180)

**AUTHORS** Strausberg, R.

**TITLE** Direct Submission

**JOURNAL** Submitted (02-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK** NIH-MGC Project URL: <http://mgc.nci.nih.gov>

## COMMENT

Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 27 Row: 1 Column: 5  
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.

## FEATURES

Location/Qualifiers

1..1180

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="LocusID:51659"

/db\_xref="taxon:9606"

/clone="MGC:19836 IMAGE:4098007"

/tissue\_type="Brain, neuroblastoma"

/clone\_lib="NIH\_MGC\_19"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

68..625

/codon\_start=1

/product="HSPC037 protein"

/protein\_id="AAH10164.1"

/db\_xref="GI:14603432"

/translation="MDAAEEVEFLAEKELVTIIPNFSLDKIYILIGDLGPFNPGLPVEV PLWLAINKQKCKLLPEWMDVEKLEMRDHERKEETFTPMPSPYMELTKLLLNH ASDNPKADEIRTLVKDMDTRIAKLRVSADSFVQQAHAHAKLDNLTLMEINTSGTFL TQALNHYKLRNLPLESTQSQDF"

## CDS

**BASE COUNT** 316 a 292 c 307 g 265 t

**ORIGIN**

Query Match 81.5%; Score 244.4; DB 9; Length 1180;

Best Local Similarity 95.1%; Pred. No. 9.2e-55;

Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 1 ATCATGCTTCAGACAACATCCGGAAGGCAGAGCAATCCGGACCTGGTCAAGGATATCT 60

|||||

DB 369 ATCATGCTTCAGACAACATCCGGAAGGCAGAGCAATCCGGACCTGGTCAAGGATATCT 428

QY 61 GGGACACTCGTATACCAACTCCGAGTGTCTGTGACAGCTTTGTGACAGCAGGAGG 120

|||||

DB 429 GGGACACTCGTATACCAACTCCGAGTGTCTGTGACAGCTTTGTGACAGCAGGAGG 488

QY 121 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACACCGCGGACTTTCCCTCA 180

|||||

DB 489 CACATGCCAAGCTGGATAACTTGACCTTGATGGAGATCAACACACCGCGGACTTTCCCTCA 548

QY 181 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTCGAAGACCTC 240

|||||

DB 549 CACAAGCGCTCAACACATGATCAAACTCCGACGAACTCCAGCTCTCGAAGACTACTC 608

QY 241 AG--CTAGGACTTCTA-AAAAAGGCGCTGGTGCAGCGCTTGGTGGG 285

|||||

DB 609 AGTCTCAGGACTTCTAGAGAAAGGCGCTGGTGCAGCGCGCTTGGTGGG 656

## RESULT 5

```

AF201939          1188 bp      mRNA      linear      PRI 20-JUL-2000
LOCUS             Homo sapiens DC5 (DC5) mRNA, complete cds.
DEFINITION        AF201939
ACCESSION         AF201939
VERSION           AF201939.1 GI:9295181
KEYWORDS          FLI CDNA.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
REFERENCE          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE             Li,Y., Peng,Y., Li,N., Gu,W., Han,Z., Fu,G. and Chen,Z.
JOURNAL           Novel genes expressed in human dentritic cell
REFERENCE          2 (bases 1 to 1188)
AUTHORS           Li,Y., Peng,Y., Li,N., Gu,W., Han,Z., Fu,G. and Chen,Z.
TITLE             Direct Submission
JOURNAL           Submitted (05-NOV-1999) Chinese National Human Genome Center at
                  Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
                  Shanghai 201203, P. R. China
FEATURES           Location/Qualifiers
                    1..1188
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="dentritic cell"
                    2..1188
                     /gene="DC5"
                     /transcript="DC5"
                     /protein_id="AAP86875.1"
                     /db_xref="GI:9295182"
                     /translation="MDAAVEFLAEKELVTIIPNFSLDKIYLLIGDLPENPGLPVEV
                     PLTAIHLKORCLLPPEWMDVEKLEKRDHERKEETFTPMSPYIMELTKILLNH
                     ASDNIPKADIRILVKDMWTRIAKLVSADSFVRQEAHAKLDNLIMEINTSGTFL
                     TQALNHMKYRTHLQPLESTQSQDF"
BASE COUNT        314 a 299 c 307 g 268 t
ORIGIN
Query Match       81.5%; Score 244.4; DB 9; Length 1188;
Best Local Similarity 95.18; Pred. No. 9.2e-55;
Matches 274; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

Qy 1 ATCATGCTTCAGACAACATCCGGAAGGACGACGAAATCCGGACCCCTGGTCAAGGATATGT 60
    |||||
Db 378 ATCATGCTTCAGACAACATCCGGAAGGACGACGAAATCCGGACCCCTGGTCAAGGATATGT 437
    |||||

Qy 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGACAGCTTTGTGACAGCAGCAGGAGG 120
    |||||
Db 438 GGGACACTCGTATAGCCAACTCCGAGTGTCTGACAGCTTTGTGACAGCAGCAGGAGG 497
    |||||

Qy 121 CACATGCCAAGCTGGATAACTTGACCTTTGATGGAGATCAACACGCGGGACTTTCTCTCA 180
    |||||
Db 498 CACATGCCAAGCTGGATAACTTGACCTTTGATGGAGATCAACACGCGGGACTTTCTCTCA 557
    |||||

Qy 181 CACAAGCGCTCAACACCATGTACAACTCCGACGAACTCCAGCCCTCTGGAAGACCTC 240
    |||||
Db 558 CACAAGCGCTCAACACCATGTACAACTCCGACGAACTCCAGCCCTCTGGAAGACCTC 617
    |||||

Qy 241 AG--CTAGGACTTCTA-AAAAGGCTTGTGACGCCCTGGTGGGG 285
    || |||||
Db 618 AGTCTCAGGACTTCTTAGAAGAAAGGCTTGTGACGCCCTGGTGGGG 665
    || |||||

RESULT 6
LOCUS             BD147616
DEFINITION        BD147616
ACCESSION         BD147616
VERSION           BD147616.1 GI:27853374
KEYWORDS          JP 2002191363-A/2459.

```

```

SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
REFERENCE          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE             Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
JOURNAL           Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
                  Patient: JP 2002191363-A 2459 09-JUL-2002;
                  HELIX RESEARCH INSTITUTE
COMMENT           OS Homo sapiens (human)
                  PN JP 2002191363-A/2459
                  PD 09-JUL-2002
                  PF 28-JUL-2000 JP 2000280990
                  PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
                  PI SAITO,
                  PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
                  PI KEIICHI NAGAI,TETSUJI OTSUKI
                  PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
                  10,
                  PC C12P21/02,C12O1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
                  Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES           Location/Qualifiers
                    1..667
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT        171 a 176 c 177 g 137 t 6 others
ORIGIN
Query Match       81.08; Score 243; DB 6; Length 667;
Best Local Similarity 94.48; Pred. No. 2.4e-54;
Matches 272; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

Qy 1 ATCATGCTTCAGACAACATCCGGAAGGACGACGAAATCCGACCCCTGGTCAAGGATATGT 60
    |||||
Db 337 ATCATGCTTCAGACAACATCCGGAAGGACGACGAAATCCGACCCCTGGTCAAGGATATGT 396
    |||||

Qy 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGACAGCTTTGTGACAGCAGCAGGAGG 120
    |||||
Db 397 GGGACACTCGTATAGCCAACTCCGAGTGTCTGACAGCTTTGTGACAGCAGCAGGAGG 456
    |||||

Qy 121 CACATGCCAAGCTGGATAACTTGACCTTTGATGGAGATCAACACGCGGGACTTTCTCTCA 180
    |||||
Db 457 CACATGCCAAGCTGGATAACTTGACCTTTGATGGAGATCAACACGCGGGACTTTCTCTCA 516
    |||||

Qy 181 CACAAGCGCTCAACACCATGTACAACTCCGACGAACTCCAGCCCTCTGGAAGACCTC 240
    |||||
Db 517 CACAAGCGCTCAACACCATGTACAACTCCGACGAACTCCAGCCCTCTGGAAGACTC 576
    |||||

Qy 241 AG--CTAGGACTTCTA-AAAAGGCTTGTGACGCCCTGGTGGGG 285
    || |||||
Db 577 AGTCTCAGGACTTCTTANANAAAGGCTTGTGACGCCCTGGTGGGG 624
    || |||||

RESULT 7
LOCUS             BC003186
DEFINITION        BC003186
ACCESSION         BC003186
VERSION           BC003186.1 GI:13112024
KEYWORDS          MGC.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
REFERENCE          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                  1 (bases 1 to 1203)
                  Strausberg,R.

```

**TITLE**  
JOURNAL

**REMARK**  
**COMMENT**

Direct Submission  
Submitted (13-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ruben Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@cgsc.bc.ca](mailto:info@cgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 6 Row: d Column: 16  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4929712.  
**FEATURES**  
source  
1..1203  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LocusID:51659"  
/db\_xref="taxon:9606"  
/clone="MGC:673 IMAGE:3503268"  
/tissue\_type="Brain, neuroblastoma"  
/clone\_lib="NIH\_MGC\_19"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
85..642  
/codon\_start=1  
/product="HSPC037 protein"  
/protein\_id="AAH03186.1"  
/db\_xref="GI:13112025"  
/translation="MDAAEVEFLAEKELVTIIPNLSLKIYLIIGDLGFPNGLPVEV  
PLWLAINKQKQKRLLPPEWMDVEKLMRDHREKEETFTPMPSYIMELTKLLNH  
ASDNIPTKADEITLVKDMMDTRIALLRVSAQSFVRQEAHAKLDNLTLMEINTSGTL  
TQALNHMYKLRTNLOPLESTQSQDF"  
BASE COUNT 326 a 300 c 309 g 268 t

**CDS**

Query Match 80.9%; Score 242.8; DB 9; Length 1203;  
Best Local Similarity 94.8%; Pred. No. 2.5e-54;  
Matches 273; Conservative 0; Mismatches 12; Indels 3; Gaps 2;  
QY 1 ATCATGCTTCAGACATCCCGAAGGCGAGAGCAATTCGGACCTGGTCAAGGATATGT 60  
|||||  
DB 386 ATCATGCTTCAGACATCCCGAAGGCGAGAGCAATTCGGACCTGGTCAAGGATATGT 445  
|||||  
QY 61 GGAGCACTCGTATAGCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAG 120  
|||||  
DB 446 GGAGCACTCGTATAGCAAACTCCGAGTGTCTGCTGACAGCTTTGTGAGACAGCAGGAG 505  
|||||  
QY 121 CACATCCCAAGCTGGTAACCTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 180  
|||||  
DB 506 CACATCCCAAGCTGGTAACCTTGACCTTGATGGAGATCAACACCGCGGACTTTCCTCA 565  
|||||  
QY 181 CACAAGCGCTCAACCATGATCAAACTCCGACAGCACTCCAGCCTCTGGAAAGACCTC 240  
|||||  
DB 566 CACAAGCGCTCAACCATGATCAAACTCCGACAGCACTCCAGCCTCTGGAGAGTACTC 625  
|||||  
QY 241 AG--CTAGGACTTCTA-AAAAGGCTGTGGTGGAGCGGCTTGGTGGG 285  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

**Db** 626 AGTCTCAGGACTTCTAGAGAAAGGCGCTGTGTGACAGCGGCTTGCTGGG 673

**RESULT** 8

**LOCUS** HS1100H13

**DEFINITION** Human DNA sequence from clone RP5-1100H13 on chromosome 20q11.2  
Contains the 3' end of gene KIAA1219, a putative novel gene, a DC5  
pseudogene, the gene for a putative RhoGAP domain containing  
protein, ESTs, STSs, GSSs and two CpG islands, complete sequence.

**ACCESSION** AL035419

**VERSION** AL035419.12 GI:11546045

**KEYWORDS** HTG: CpG island; DC5; KIAA1219; RhoGAP.

**SOURCE** Homo sapiens (human)

**ORGANISM** Homo sapiens

**REFERENCE** 1 (bases 1 to 116792)

**AUTHORS** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**TITLE** Direct Submission

**JOURNAL** Submitted (12-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

**COMMENT** On Dec 5, 2000 this sequence version replaced gi:10198625.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr20>  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP5-1100H13 is from  
the library RPCI-5 constructed by the group of Pieter de Jong. For  
further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pCIPAC2  
This sequence is the entire insert of clone RP5-1100H13 The true  
left end of clone RP11-12201 is at 47827 in this sequence. The true  
right end of clone RP5-927M24 is at 16404 in this sequence.

**FEATURES**  
source  
1..116792  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="RZPD:RPCIP704H131100"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="q11.2"  
/clone="RP5-1100H13"  
/clone\_lib="RPCI-5"  
1331..1465  
/note="MER58C repeat: matches 68..89 of consensus"  
1587..13349  
/gene="djl100H13.1"  
join(<1587..1723,4379..4487,5260..5338,8641..8789,  
9265..13349)  
/gene="djl100H13.1"  
/product="djl100H13.1.1 (KIAA1219 (similar to Drosophila  
H09358 and C. elegans D2085.5) isoform 1)."  
/note="match: cDNAs: Em:AB033045 Em:AL132998

**repeat\_region**

**gene**

**mRNA**

```

match: ESTs: Em: A1190579 Em: AA256927 Em: N36561 Em: AA874351
Em: A1281190 Em: AA180094 Em: AA669533 Em: AA511971
Em: A1195307 Em: AA215757 Em: AA156777 Em: AA131113
Em: AA418776 Em: A1357456 Em: A1379408 Em: N76170 Em: W52362
Em: AA024901 Em: A1334294 Em: AA205264 Em: AA215603
Em: AA418686 Em: AA024902 Em: A1379162 Em: AA424825
Em: A1202970 Em: A1248059 Em: AA648752 Em: A1359832
Em: A1366165 Em: A12224862 Em: A1359759 Em: A1492050 Em: R76319
Em: A1050034 Em: A1423927 Em: AA024896 Em: R78702 Em: A1066457
Em: R72124 Em: A1361138 Em: R81299 Em: AA282907 Em: L44403
Em: A1500502 Em: Z98499 Em: R83793 Em: AA412128 Em: AA179405
Em: AA459722 Em: AA173335 Em: AA283091 Em: H26506 Em: H02652
Em: AA258043 Em: N64761 Em: AA255436 Em: A1271673 Em: A1017681
Em: AA459863 Em: A1221344 Em: R65916 Em: W07865 Em: AA383163
Em: H26520 Em: AA872554 Em: AA343687 Em: AA889796 Em: R68642
Em: AA580081 Em: AA156671 Em: AA552080 Em: AA024895"
/evidence-not_experimental
join(<1587, .1723, 4379, .4487, 5260, .5338, 8641, .8789,
9265, .9458)
/gene="dJ1100H13.1"
/note="Continues in Em: AL049868 as dJ927M24.2
match: proteins: Tr: Q9Y3G9 Tr: Q9W2X2 Tr: Q9U4G6 Tr: Q18431"
/codon_start=3
/evidence-not_experimental
/product="dJ1100H13.1.1 (K1AA1219 (similar to Drosophila
GH03358 and C. elegans D2085.5) isoform 1)"
/protein_id="CABA1646.1"
/db_xref="GI: 4704208"
/db_xref="SPTREMBL: Q9Y3G9"
/translation="DSLESNTSDSDSNMDLMPILKOPSLTELPNHTDNLSSQ
RLSPSSRMRLPOGRPPVPLGPETRVSVWVDDIENPLSLMTEISGVEYTTAN
SSTSRLSTLEKEVPFIHPLNGLFRKIQATGAFNMVPLVDGMIVSRRALGEL
VRQVVICRRKRLESDSYSPHVRKQKITDYNKYRNKQLEPEFTYLSLQEVGLKN
CSS"
repeat_region 2753, .3187
/note="L1M4 repeat: matches 3867, .4299 of consensus"
4727, .5189
/gene="dJ1100H13.1"
/note="match: GSS: Em: A0663797"
4732, .4875
/note="match: GSS: Em: A0663797"
/gene="dJ1100H13.1"
/note="match: STS: Em: HSPF04E1"
4745, .5581
/gene="dJ1100H13.1"
/note="match: GSS: Em: A0745415"
complement(4997, .5442)
/note="match: STS: Em: G33216"
complement(5210, .5903)
/note="match: GSS: Em: A0321022"
5522, .6464
/note="L1M40 repeat: matches 5388, .6322 of consensus"
6837, .6970
/note="MIR repeat: matches 32, .175 of consensus"
complement(6984, .7365)
/note="match: GSS: Em: A0807920"
7232, .7391
/note="L1M4 repeat: matches 7614, .7783 of consensus"
7610, .7684
/note="AluJb repeat: matches 39, .113 of consensus"
join(<8150, .8789, 9265, .9431, 10364, .>10484)
/gene="dJ1100H13.1"
/product="dJ1100H13.1.2 (K1AA1219, isoform 2)"
/note="match: ESTs: Em: A1536595"
/evidence-not_experimental
8567, .8616
/note="2 copies 25 mer 98% conserved"
12187, .12234
/note="12 copies 4 mer caca 89% conserved"
12536, .12571
/note="18 copies 2 mer aa 80% conserved"
12911, .12955
/note="L2 repeat: matches 2657, .2701 of consensus"
complement(12943, .13346)

```

CDS

```

polyA_signal
polyA_site
repeat_region
gene
mRNA
mRNA
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
polyA_site
mRNA
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

```

```

Query Match 72.1%; Score 216.4; DB 9; Length 116792;
Best Local Similarity 91.3%; Pred. No. 1.6e-47;
Matches 263; Conservative 0; Mismatches 21; Indels 4; Gaps 3;

QY 1 ATCATGCTTCAGACAACTCCGAGGACAGCAAGAAATCCGACCTGGTCAAGATATGT 60
Db 49659 ATCATGCTTCAGACAACTCCGAGGACAGCAAGAAATCCGACCTGGTCAAGATAT 49600
QY 61 GGGACACTCGTATAGCCAAACTCGAGTGTCTGTGACAGTTTGTGACAGCAGCAGG 120
Db 49599 GGGACACTCGTATAGCCAAACTCGAGTGTCTGTGACAGTTTGTGACAGCAGG 49540
QY 121 CACATGCCAAGCTGGATAACTTGTACCTTGTATGGAGATCAACACCGCGGACTTTCCTCA 180

```

```

||||| 49539 CACATGCCAAGCCGATACCTTACCTTATGGATGATCAACCCCGAGGACTTTCCTCA 49480
||||| 181 CACAAGCGCTCAACACATGTAACAACTCCGACGAACTCCAGCCTCTGGAAGACCTC 240
||||| 49479 CACAAGCGCTCAACACATGTAACAACTCCGACGAACTCCAGCCTCTGGAAGACCTC 49421
||||| 241 AG--CTAGACACTCTA-AAAAAGGCTGTGTGACGCGCTGTGTGGG 285
||||| 49420 AGTCTCAGGACTCTAGAGAAAGGCTGTGTGACGAGCTTGTCTGGG 49373

RESULT 9
AX336730
LOCUS AX336730 454 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 7239 from Patent WO0194629.
ACCESSION AX336730
VERSION AX336730.1 GI:18127449
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrihan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 7239 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..454
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 128 a 113 c 116 g 97 t
ORIGIN
Query Match 52.7%; Score 158; DB 6; Length 454;
Best Local Similarity 98.9%; Pred. No. 1.5e-31;
Matches 180; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 ATCATGCTTCAGACAACTCCGAGGAGAGAGAAATCCGGACCTTGTCAAGGATATGT 60
||||| 274 ATCATGCTTCAGACAACTCCGAA-GCAGAGAAATCCGGACCTTGTCAAGGATATGT 332
||||| 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGTGACAGCTTGTGACAGCAGGAGG 120
||||| 333 GGGACACTCGTATAGCCAACTCCGAGTGTCTGTGACAGCTTGTGACAGCAGGAGG 392
||||| 121 CACATGCCAAGCTGATGACTTGTGATGAGATCAACACCGCGGACTTTCCTCA 180
||||| 393 CACATG-CAAGCTGGATACCTTGTGATGAGATCAACACCGCGGACTTTCCTCA 451
QY 181 CA 182
Db 452 CA 453

RESULT 10
AB097169
LOCUS AB097169 933 bp DNA linear VRT 09-MAY-2003
DEFINITION Xenopus laevis Psf2 gene, complete cds.
ACCESSION AB097169
VERSION AB097169.1 GI:29365480
KEYWORDS Xenopus laevis (African clawed frog)
SOURCE Xenopus laevis
ORGANISM Xenopus laevis
REFERENCE
AUTHORS Kubota,Y., Takase,Y., Komori,Y., Hashimoto,Y., Arata,T.,

```

# TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL

Kamimura,Y., Araki,H. and Takisawa,H.  
A novel ring-like complex of Xenopus proteins essential for the  
initiation of DNA replication  
Genes Dev. 17 (9), 1141-1152 (2003)  
22615695  
12730133  
2 (bases 1 to 933)  
Kubota,Y. and Takisawa,H.  
Direct Submission  
Submitted (28-NOV-2002) Yumiko Kubota, Osaka University, Department  
of Biology, Graduate School of Science, Machikaneyamacho 1-1,  
Toyonaka, Osaka 560-0043, Japan  
(E-mail:ykubota@bio.sci.osaka-u.ac.jp, Tel:81-6-6850-5554,  
Fax:81-6-6850-5554)  
Location/Qualifiers  
1..933  
/organism="Xenopus laevis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:8355"  
21..578  
/gene="Psf2"  
21..578  
/gene="Psf2"  
/function="essential for DNA replication"  
/codon\_start=1  
/product="Psf2"  
/protein\_id="BAC66459.1"  
/db\_xref="GI:29365481"  
/translation="MDASEVEFLAEKEQVTVIPNSLDKVLKYLGGDLGPNFSLPVEV  
PLWLAINLQKQKCRIVPPEDWDEKLEAIRDOERBEETFTPMSPYYMLKLLLNH  
AADNIPKADEIFTLVKDWTDTRIAKLRSADSFVRQEAHAKLDNLTLMEINTIGTFF  
TESLNHYKLRTSLQNPERSQSDY"  
BASE COUNT 282 a 200 c 208 g 243 t  
ORIGIN

## FEATURES source

## gene CDS

Query Match 45.5%; Score 136.4; DB 5; Length 933;  
Best Local Similarity 72.7%; Pred. No. 8.1e-26;  
Matches 176; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1 ATCATGCTTCAGACAACTCCGAGGAGAGAGAAATCCGGACCTTGTCAAGGATATGT 60  
||||| 322 ATCATGCTTCAGACAACTCCCTTAAGCAGATGAGATTCGTACACTTGTAAAGACACCT 381  
QY 61 GGGACACTCGTATAGCCAACTCCGAGTGTCTGTGACAGCTTGTGACAGCAGCAGGAG 120  
||||| 382 GGGATACAAGATAGCAAACTCGGCTCTCTGTGACAGCTTGTGAAGGGCAGGAAG 441  
QY 121 CACATGCCAAGCTGATGACTTGTGATGAGATCAACACCGCGGACTTTCCTCA 180  
||||| 442 CTGATGCCAAGCTGGATAACCTGACGCTAATGGAAATTAACACCATCGGAACATTTT 501  
QY 181 CACAAGCGCTCAACACATGTACAACTCCGACGAACTCCAGCCTCTGGAAGACCTC 240  
||||| 502 CTGAGTCTTTAAACCCACATGTACAGCTACCTACCTGACAGACCCAGAGAGAGGAC 561  
QY 241 AG 242  
Db 562 AG 563

## RESULT 11 AC009068/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS

AC009068 160356 bp DNA linear PRI 27-APR-2001  
Homo sapiens chromosome 16 clone RP11-314K3, complete sequence.  
AC009068  
AC009068.10 GI:13811892  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 160356)  
DOE Joint Genome Institute and Stanford Human Genome Center.











- \* is not known and their order in this sequence record is
- \* arbitrary. Gaps between the contigs are represented as
- \* runs of N, but the exact sizes of the gaps are unknown.
- \* This record will be updated with the finished sequence.
- \* as soon as it is available and the accession number will
- \* be preserved.

2685:	contig of 2685 bp in length
2785:	gap of 100 bp
3775:	contig of 990 bp in length
3875:	gap of 100 bp
112935:	contig of 103060 bp in length
113035:	gap of 100 bp
115741:	contig of 2706 bp in length
115841:	gap of 100 bp
130109:	contig of 14268 bp in length
130209:	gap of 100 bp
138730:	contig of 8521 bp in length
138830:	gap of 100 bp
186703:	contig of 47873 bp in length
186803:	gap of 100 bp
188292:	contig of 1489 bp in length.

## FEATURES

```

1. .188292
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-41A22"
/clone_lib="RPCI-22 Male Mouse BAC"

```

```
misc_feature      1. 2685
```

```
1. 2003
/note="assembly_fragment
clone_end:SP6
vector_side:left"
```

misc\_feature 2786. :3775

```

/note="assembly_fragment"
3876. 112935

```

```
=====
06/07 :112559
/note="assembly fragment"
```

```

/proc= assembly_115741
113036 115741

```

```

"msc_reduce
115050: .115/41
/not-a-assembly fragment"

```

```

115842 130100
/note= assembly_fragment

```

II384Z: :I30I09  
/no+o=7-ccomb1::  
f-ccomb1

```
130210 130720
/note=assembly_fragment-
```

130210: .138730

```

1200001 /note="assembly_fragment"
1200002
1200003

```

138831. .186703

```

/note="assembly_fragment"

```

186804. .188292

```
/note="assembly_fragment
```

clone\_end:T7

BASE COUNT	vector_side:right"	
43856 a	49637 c	49457 g
ORIGIN		44640 t
		702 others

	Query Match	33.2%	Score 99.6;	DB 2;	Length 188292;
	Best Local Similarity	72.5%;	Pred. No. 2.9e-16;		
	Matches 129;	Conservative 0;	Mismatches 49;	Indels 0;	Gaps 0;
QY	5	TGCTTCAGACAACATCCCGAAGGCGAGACGAAATCCGGACCTCTGGTCAAGGATATCTGGGA	64		
Db	55088	TGCTTCTGACACATCCCCAAGCAGACACACATCCGGACATGATCAAGACCTGTGGGA	55029		
QY	65	CACCTCGTATAGCCAAACTCCGAGTGTCTGCTGACAGCTTTGTGACACAGCAGGAGGCACA	124		
Db	55028	CACACGCATGGCCAAAGCTTCGAGTGTCTGCTGACAGTTTGTGCGCGCAGCAGAGGCACA	54969		
QY	125	TGCCAAGCTGGATAACTTGCACCTTATGGAGATCAACACACGCGGGACTTTCCTCACA	182		
Db	54968	TGCCAAGGTAGGCATGGTTTCTCTGTCCCTGTCGCGGTGCCAGTGGGCACTGCCTGCCA	54911		

Search completed: October 14, 2003, 17:02:31  
Job time : 1479 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 11:32:16 ; Search time 1272.67 Seconds  
(without alignments)  
5729.185 Million cell updates/sec

Title: US-09-854-124-6  
Perfect score: 300  
Sequence: 1 aattccgtgctgctggtga.....cgaaaggaagaactttac 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_esti:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	286.4	95.5	959	10	BG118214
2	285	95.0	714	10	BG104289
3	284.8	94.9	388	9	AI914017
4	284.8	94.9	429	10	BE262182
					-----
					BG118214 602350003
					BG104289 602310878
					AI914017 wd74803.x
					BE262182 601150452

5	284.8	94.9	487	9	AI827298
6	284.8	94.9	488	9	AA251319
7	284.8	94.9	515	10	BF312015
8	284.8	94.9	532	9	AI583174
9	284.8	94.9	550	9	AI828992
10	284.8	94.9	573	10	BE514578
11	284.8	94.9	671	10	BE258240
12	284.8	94.9	696	9	AW249012
13	284.8	94.9	702	9	AV712739
14	284.8	94.9	716	12	BG825252
15	284.8	94.9	729	10	BE251621
16	284.8	94.9	731	10	BG475509
17	284.8	94.9	736	10	BE312319
18	284.8	94.9	740	13	BU595469
19	284.8	94.9	749	12	BG772776
20	284.8	94.9	761	10	BG339264
21	284.8	94.9	761	12	BI257993
22	284.8	94.9	771	10	BG475384
23	284.8	94.9	833	13	BU597296
24	284.8	94.9	834	10	BG527542
25	284.8	94.9	877	13	BU940719
26	284.8	94.9	882	10	BF684568
27	284.8	94.9	891	11	BC022839
28	284.8	94.9	897	10	BG519560
29	284.8	94.9	905	9	AI525760
30	284.8	94.9	909	13	BQ229290
31	284.8	94.9	914	14	CA455226
32	284.8	94.9	943	10	BF683514
33	284.8	94.9	982	13	BU957713
34	284.8	94.9	997	10	BG420251
35	284.8	94.9	1007	10	BF795838
36	284.8	94.9	1030	10	BF311745
37	284.8	94.9	1057	10	BF312439
38	284.8	94.9	1067	13	BX415204
39	284.8	94.9	1182	11	AF125098
40	284.8	94.9	1193	9	AL529785
41	284	94.7	624	10	BG284180
42	284	94.7	908	10	BE267221
43	284	94.7	1201	9	AL560880
44	283.2	94.4	568	10	AW958853
45	283.2	94.4	683	10	BE260083

#### ALIGNMENTS

```

RESULT 1
BG118214
LOCUS      BG118214
DEFINITION 602350003f1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:444801 5',
            mRNA sequence.
ACCESSION  BG118214
VERSION     BG118214.1
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 959)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAMI0221 row: f column: 02
            High quality sequence stop: 719.

```

BG118214 602350003f1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:444801 5',  
mRNA sequence.

ACCESSION BG118214  
VERSION BG118214.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 959)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAMI0221 row: f column: 02

High quality sequence stop: 719.

FEATURES  
source

Location/Qualifiers  
1. 959  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:444801"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_90"  
/note="Organ: liver; Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC library." 1 others

BASE COUNT 249 a 245 c 273 g 191 t  
ORIGIN

Query Match 95.5%; Score 286.4; DB 10; Length 959;  
Best Local Similarity 98.0%; Pred. No. 7.9e-67;  
Matches 290; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 5 CCGTTGCTCGGTGAGGCTCTGGCCCTCAGCTCGCGCCGCGCATGGAGCTGCCGAGGTC 64  
||| |  
Db 29 CCGAGGACCGGTGAGGCTCTGGCCCTCAGCTCGCGCCGCGCATGGAGCTGCCGAGGTC 88  
||| |  
QY 65 GAATTCCTCGCGGAGGAGCTGTTACCAATTATCCCAACTTCAGTCTGGACAAGATC 124  
||| |  
Db 89 GAATTCCTCGCGGAGGAGCTGTTACCAATTATCCCAACTTCAGTCTGGACAAGATC 148  
||| |  
QY 125 TACCTCATCGGGGGGACCTGGGGCTTTTAACTGTTTACCGTGGAGAGTCCCTG 184  
||| |  
Db 149 TACCTCATCGGGGGGACCTGGGGCTTTTAACTGTTTACCGTGGAGAGTCCCTG 208  
||| |  
QY 195 TGGCTGGCGATTAACTGAAACAAAGACAGAAATGTGCCCTGCTCCCTCCAGATGGATG 244  
||| |  
Db 209 TGGCTGGCGATTAACTGAAACAAAGACAGAAATGTGCCCTGCTCCCTCCAGATGGATG 268  
||| |  
QY 245 GATGTAGAAAAGTTGGAGAGATGAGGATCATGACGAAAGGAGAACTTTTAC 300  
||| |  
Db 269 GATGTAGAAAAGTTGGAGAGATGAGGATCATGACGAAAGGAGAACTTTTAC 324  
||| |

RESULT 2  
BG104289 714 bp mRNA linear EST 30-JAN-2001  
LOCUS 602310878F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:4423051 5',  
DEFINITION mRNA sequence.  
ACCESSION BG104289  
VERSION BG104289  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 714)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LLCMI218 row: k column: 20  
High quality sequence stop: 702.  
Location/Qualifiers  
1. 714  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

FEATURES  
source

Location/Qualifiers  
1. 388  
/organism="Homo sapiens"

/clone="IMAGE:4423051"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_20"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 186 a 182 c 196 g 150 t  
ORIGIN

Query Match 95.0%; Score 285; DB 10; Length 714;  
Best Local Similarity 100.0%; Pred. No. 1.8e-66;  
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GGTGAGGCTCTGGGCTCGAGCTCGCGCGCGCATGGAGTTCGAATTCCTCGC 75  
||| |  
Db 1 GGTGAGGCTCTGGGCTCGAGCTCGCGCGCGCATGGAGTTCGAATTCCTCGC 60  
||| |  
QY 76 CGAAGAGAGCTGTTTACCAATTATCCCAACTTCAGTCTGGACAAGATCTACCTCATCGG 135  
||| |  
Db 61 CGAAGAGAGCTGTTTACCAATTATCCCAACTTCAGTCTGGACAAGATCTACCTCATCGG 120  
||| |  
QY 136 GGGGACCTGGGGCTTTTAACTGTTTACCGTGGAGTTCGCTGCTCCAGATGGATGTAGAAA 195  
||| |  
Db 121 GGGGACCTGGGGCTTTTAACTGTTTACCGTGGAGTTCGCTGCTCCAGATGGATGTAGAAA 180  
||| |  
QY 196 TAACTGAAACAAAGACAGAAATGTGCCCTGCTCCCTCCAGATGGATGTAGAAA 255  
||| |  
Db 181 TAACTGAAACAAAGACAGAAATGTGCCCTGCTCCCTCCAGATGGATGTAGAAA 240  
||| |  
QY 256 GTTGGAGAGATGAGGATCATGACGAAAGGAGAACTTTTAC 300  
||| |  
Db 241 GTTGGAGAGATGAGGATCATGACGAAAGGAGAACTTTTAC 285  
||| |

RESULT 3  
AT914017 388 bp mRNA linear EST 17-DEC-1999  
LOCUS wd74a03.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2337292 3',  
DEFINITION similar to WP:F31C3.5 CE17734 i, mRNA sequence.  
ACCESSION AT914017  
VERSION AT914017  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 388)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
JOURNAL Tumor Gene Index  
COMMENT Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone Distribution: Washington University Genome Sequencing Center  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert length: 1146 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 384.  
Location/Qualifiers  
1. 388  
/organism="Homo sapiens"



/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2422723"  
 /tissue\_type="squamous cell carcinoma, poorly  
 differentiated (4 pooled tumors, including primary and  
 metastatic)"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NCI\_CGAP\_Lu19"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from a  
 pooled lung tumor tissue, and was then primed with a Not I  
 - oligo(dT) primer. Double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT73 vector. Library went through one round of  
 normalization. Library constructed by Bento Soares and M.  
 Fatima Bonaldo."

BASE COUNT 119 a 132 c 138 g 98 t

Query Match 94.9%; Score 284.8; DB 9; Length 487;

Best Local Similarity 99.3%; Pred. No. 1.8e-66;

Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 GTCGGTGGAGGCTGTGGCTGCAGCTGCGCGCCGCGCATGGAGCTGCGAGGTGGAATTCCT 72

Db 43 GACCGTGGAGGCTGTGGCTGCAGCTGCGCGCCGCGCATGGAGCTGCGAGGTGGAATTCCT 102

QY 73 CGCCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 132

Db 103 CGCCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 162

QY 133 CGGGGGGACCTGGGGGCTTTAAACCTGCTTTACCCGCGAGTGGCCCTGTGGCTGCG 192

Db 163 CGGGGGGACCTGGGGGCTTTAAACCTGCTTTACCCGCGAGTGGCCCTGTGGCTGCG 222

QY 193 GATTAACCTGAAACAAGACAGAAATGTCCCTGCTCTCCAGAGTGGATGGATGAGA 252

Db 223 GATTAACCTGAAACAAGACAGAAATGTCCCTGCTCTCCAGAGTGGATGGATGAGA 282

QY 253 AAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 300

Db 283 AAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 330

## RESULT 6

AA251319

LOCUS  
 DEFINITION

zs04d09.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:684209 3',  
 similar to SW:YJH2\_YEAST P40359 HYPOPHETICAL 25.1 KD PROTEIN IN  
 SCPI60-MRPL8 INTERGENIC REGION. ;, mRNA sequence.

ACCESSION  
 VERSION

AA251319  
 AA251319.1 GI:1886282

EST.  
 Homo sapiens (human)

ORGANISM  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 488)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Possible reversed clone: polyT not found  
 Insert length: 910 Std Error: 0.00  
 Seq primer: -41ml3 fwd. ET from Amersham

JOURNAL  
 COMMENT

## FEATURES

source

High quality sequence stop: 463.

Location/Qualifiers

1. 488

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:684209"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_GCB1"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand  
 cDNA was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-GTTACCAATCTGAAGTGGGCGCCGCTCATTTTTTTTTTTTTTT-3'  
 ]. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 118 a 134 c 138 g 98 t

Query Match 94.9%; Score 284.8; DB 9; Length 488;

Best Local Similarity 99.3%; Pred. No. 1.8e-66;

Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 GTCGGTGGAGGCTGTGGCTGCAGCTGCGCGCCGCGCATGGAGCTGCGAGGTGGAATTCCT 72

Db 43 GACCGTGGAGGCTGTGGCTGCAGCTGCGCGCCGCGCATGGAGCTGCGAGGTGGAATTCCT 102

QY 73 CGCCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 132

Db 103 CGCCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCAT 162

QY 133 CGGGGGGACCTGGGGGCTTTAAACCTGCTTTACCCGCGAGTGGCCCTGTGGCTGCG 192

Db 163 CGGGGGGACCTGGGGGCTTTAAACCTGCTTTACCCGCGAGTGGCCCTGTGGCTGCG 222

QY 193 GATTAACCTGAAACAAGACAGAAATGTCCCTGCTCTCCAGAGTGGATGGATGAGA 252

Db 223 GATTAACCTGAAACAAGACAGAAATGTCCCTGCTCTCCAGAGTGGATGGATGAGA 282

QY 253 AAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 300

Db 283 AAGTTGGAGAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 330

## RESULT 7

BF312015

LOCUS  
 DEFINITION

601897874F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:412178 5',  
 mRNA sequence.

ACCESSION  
 VERSION

BF312015  
 BF312015.1 GI:11259802

EST.  
 Homo sapiens (human)

ORGANISM  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 515)

NIH-MGC http://mgc.ncbi.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

Tissue procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.

JOURNAL  
 COMMENT



Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLCMI018 row: c column: 19  
 High quality sequence stop: 515.  
 Location/Qualifiers  
 1. 515

#### FEATURES

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4127178"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_19"
/notes="Organ: brain; Vector: POTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

```

```

BASE COUNT      132 a  134 c  146 g  103 t
ORIGIN
Query Match      94.9%; Score 284.8; DB 10; Length 515;
Best Local Similarity 99.3%; Pred. No. 1.9e-66;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 GTCGTGAGGCTCTGGCCCTGACGCTCGCGCCGCCATGGACGCTGCGAGGTGCAATTCTCT 72
DB 26 GACCGTAGGCTCTGGCCCTGACGCTCGCGCCGCCATGGACGCTGCGAGGTGCAATTCTCT 85

QY 73 CCGCGAGAAGGAGCTGGTTACCATTCATCCCAACTTCAGTCTGGACAGATCTACTCAT 132
DB 86 CGCCGAGAAGGAGCTGGTTACCATTCATCCCAACTTCAGTCTGGACAGATCTACTCAT 145

QY 133 CGGGGGGACCTGGGGCCCTTTTAACTCTGGTTTACCCCTGGTGAAGTGCCTCTGGCTGGC 192
DB 146 CGGGGGGACCTGGGGCCCTTTTAACTCTGGTTTACCCCTGGTGAAGTGCCTCTGGCTGGC 205

QY 193 GATTAACTTGAACAAAGACAGAAATGTCGCCTGCTCCCTCCAGAGTGGATGATGTAGA 252
DB 206 GATTAACTTGAACAAAGACAGAAATGTCGCCTGCTCCCTCCAGAGTGGATGATGTAGA 265

QY 253 AAAGTTGGAGAAGATGAGGAGTATCAACGAAGGAAGAAACTTTTAC 300
DB 266 AAAGTTGGAGAAGATGAGGAGTATCAACGAAGGAAGAAACTTTTAC 313

```

#### RESULT 8

```

AI583174
LOCUS      532 bp  mRNA  linear  EST 14-DEC-1999
DEFINITION t64e05.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2213600 3' similar to SW:YJH2_YEAST P40359 HYPOPHETICAL 25.1 KD PROTEIN IN SMC3-MRPL8 INTERGENIC REGION. [1]; mRNA sequence.

```

AI583174

AI583174.1 GI:4569071

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 532)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/dbrrp/image/image.html](http://www-bio.llnl.gov/dbrrp/image/image.html)  
 Insert Length: 1092 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 459  
 POLYA=No.

#### FEATURES

source

```

Location/Qualifiers
1. 532
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2213600"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/lab_clone_lib="NCI_CGAP_Lu19"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

```

```

BASE COUNT      138 a  140 c  149 g  104 t  1 others
ORIGIN
Query Match      94.9%; Score 284.8; DB 9; Length 532;
Best Local Similarity 99.3%; Pred. No. 1.9e-66;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 13 GTCGTGAGGCTCTGGCCCTGACGCTCGCGCCGCCATGGACGCTGCGAGGTGCAATTCTCT 72
DB 25 GACCGTAGGCTCTGGCCCTGACGCTCGCGCCGCCATGGACGCTGCGAGGTGCAATTCTCT 84

QY 73 CCGCGAGAAGGAGCTGGTTACCATTCATCCCAACTTCAGTCTGGACAGATCTACTCAT 132
DB 85 CCGCGAGAAGGAGCTGGTTACCATTCATCCCAACTTCAGTCTGGACAGATCTACTCAT 144

QY 133 CGGGGGGACCTGGGGCCCTTTTAACTCTGGTTTACCCCTGGTGAAGTGCCTCTGGCTGGC 192
DB 145 CGGGGGGACCTGGGGCCCTTTTAACTCTGGTTTACCCCTGGTGAAGTGCCTCTGGCTGGC 204

QY 193 GATTAACTTGAACAAAGACAGAAATGTCGCCTGCTCCCTCCAGAGTGGATGATGTAGA 252
DB 205 GATTAACTTGAACAAAGACAGAAATGTCGCCTGCTCCCTCCAGAGTGGATGATGTAGA 264

QY 253 AAAGTTGGAGAAGATGAGGAGTATCAACGAAGGAAGAAACTTTTAC 300
DB 265 AAAGTTGGAGAAGATGAGGAGTATCAACGAAGGAAGAAACTTTTAC 312

```

#### RESULT 9

AI828992

LOCUS

DEFINITION

AI828992

AI828992

AI828992.1

GI:5449663

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 550)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index



REFERENCE 1 (bases 1 to 671)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)  
 Plate: L16M161 row: i column: 24  
 High quality sequence stop: 655.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..671  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3356255"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_16"  
 /note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."  
 BASE COUNT 172 a 179 c 184 g 136 t  
 ORIGIN  
 Query Match 94.9%; Score 284.8; DB 10; Length 671;  
 Best Local Similarity 99.3%; Pred. No. 2e-66;  
 Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 13 GTCGGTGAGGCTCTGGCTGCAGCTCGCGCCCGCCATGACGCTGCCGAGGTGCAATTCCT 72  
 Db 8 GACCTTGAGGCTCTGGCTGCAGCTCGCGCCCGCCATGACGCTGCCGAGGTGCAATTCCT 67  
 QY 73 CGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAGATCTACCTCAT 132  
 Db 68 CGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAGATCTACCTCAT 127  
 QY 133 CGGGGGGACCTGGGGCCCTTTTAAACCCCTGGTTTACCCCTGGGAAGTGCCTCTGGCTGGC 192  
 Db 128 CGGGGGGACCTGGGGCCCTTTTAAACCCCTGGTTTACCCCTGGGAAGTGCCTCTGGCTGGC 187  
 QY 193 GATTAACCTGAACAAAGACAGCAAAATGTCGCTGCCCTCCAGAGTGGATGGATGAGA 252  
 Db 188 GATTAACCTGAACAAAGACAGCAAAATGTCGCTGCCCTCCAGAGTGGATGGATGAGA 247  
 QY 253 AAAGTTGGAGAAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 300  
 Db 248 AAAGTTGGAGAAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 295  
 RESULT 12  
 AW249012  
 LOCUS AW249012  
 DEFINITION 2821094.5prime NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:2821094 5', mRNA sequence.  
 ACCESSION AW249012  
 VERSION AW249012.1 GI:6592005  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 696)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Other ESTs: 2821094.3prime  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing project clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)  
 Scores: PHRED from University of Washington Genome Center. Vector Trimming: cross\_match from University of Washington Genome Center. PHRAP suite. Poly-T Identification: patmatch.pl from Berkeley Drosophila Genome Project. University of Washington Genome Center. <http://www.genome.washington.edu>  
 Plate: L16M161 row: O column: 15  
 High quality sequence stop: 573.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..696  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2821094"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_7"  
 /note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 179 a 183 c 192 g 142 t  
 ORIGIN  
 Query Match 94.9%; Score 284.8; DB 9; Length 696;  
 Best Local Similarity 99.3%; Pred. No. 2e-66;  
 Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 13 GTCGGTGAGGCTCTGGCTGCAGCTCGCGCCCGCCATGACGCTGCCGAGGTGCAATTCCT 72  
 Db 13 GACCTTGAGGCTCTGGCTGCAGCTCGCGCCCGCCATGACGCTGCCGAGGTGCAATTCCT 72  
 QY 73 CGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAGATCTACCTCAT 132  
 Db 73 CGCCGAGAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAGATCTACCTCAT 132  
 QY 133 CGGGGGGACCTGGGGCCCTTTTAAACCCCTGGTTTACCCCTGGGAAGTGCCTCTGGCTGGC 192  
 Db 133 CGGGGGGACCTGGGGCCCTTTTAAACCCCTGGTTTACCCCTGGGAAGTGCCTCTGGCTGGC 192  
 QY 193 GATTAACCTGAACAAAGACAGCAAAATGTCGCTGCCCTCCAGAGTGGATGGATGAGA 252  
 Db 193 GATTAACCTGAACAAAGACAGCAAAATGTCGCTGCCCTCCAGAGTGGATGGATGAGA 252  
 QY 253 AAAGTTGGAGAAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 300  
 Db 253 AAAGTTGGAGAAGATGAGGATCATGAACGAAGGAAGAACTTTTAC 300  
 RESULT 13  
 AV712739  
 LOCUS AV712739  
 DEFINITION AV712739 DCA Homo sapiens cDNA clone DCAACE07 5', mRNA sequence.  
 ACCESSION AV712739  
 VERSION AV712739.1 GI:10732045  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

# REFERENCE AUTHORS

1 (bases 1 to 702)  
Xu, X., Gu, J., Liu, F., Qu, J., Zhao, M., Li, Y., Huang, Q., Zhou, J.,  
Song, H., Gu, Y., Yang, Y., Gao, G., Xiao, H., Li, N., Qian, B., Gao, X.,  
Cheng, Z., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu,  
G., Cheng, Z., and Han, Z.

# TITLE JOURNAL COMMENT

Homo sapiens cDNA clones  
Unpublished  
Contact: Zequan Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chc.sh.cn  
This clone is available at CHGC in Shanghai.

# FEATURES source

Location/Qualifiers  
1..702  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DCAACE07"  
/cell\_type="dendritic cells"  
/dev\_stage="mature"  
/lab\_host="BM25.8"  
/clone\_lib="DCA"  
/note="Vector: pTriplex2; Site\_1: sfIIA; Site\_2: sfIIB"  
BASE COUNT 183 a 181 c 184 g 147 t 7 others  
ORIGIN

Query Match 94.9%; Score 284.8; DB 9; Length 702;  
Best Local Similarity 99.3%; Pred. No. 2e-66;  
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 13 GTCGGTAGGCTCGGCTCGAGCTCGCGCGCCATGAGCGTCCGAGTGCAGATTCCT 72  
Db 5 GACCGTAGGCTCGGCTCGAGCTCGCGCGCCATGAGCGTCCGAGTGCAGATTCCT 64  
Qy 73 CGCCGAGAAGAGCTGTTACCATTTATCCCAACTTCAGTCTGACAAAGATCTACCTCAT 132  
Db 65 CGCCGAGAAGAGCTGTTACCATTTATCCCAACTTCAGTCTGACAAAGATCTACCTCAT 124  
Qy 133 CGGGGGGAGCTGGGGCTTTTAACTTACCTGTTTACCGTGGAGTGCCTCGGCGGC 192  
Db 125 CGGGGGGAGCTGGGGCTTTTAACTTACCTGTTTACCGTGGAGTGCCTCGGCGGC 184  
Qy 193 GATTAACTTGAACAAAGACAGAAATGTCGCTGCTCCCTCCAGTGGATGTAGTA 252  
Db 185 GATTAACTTGAACAAAGACAGAAATGTCGCTGCTCCCTCCAGTGGATGTAGTA 244  
Qy 253 AAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAAGAACTTTTAC 300  
Db 245 AAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAAGAACTTTTAC 292

RESULT 14  
BG825252  
LOCUS 716 bp mRNA linear EST 22-MAY-2001  
DEFINITION 602747675F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4900602 5',  
mRNA sequence.

ACCESSION BG825252  
VERSION BG825252.1 GI:14172839  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 716)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLCMI794 row: e column: 19

High quality sequence stop: 716.

# FEATURES source

Location/Qualifiers  
1..716  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4900602"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_17"  
/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 186 a 185 c 198 g 147 t  
ORIGIN

Query Match 94.9%; Score 284.8; DB 12; Length 716;  
Best Local Similarity 99.3%; Pred. No. 2e-66;  
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 13 GTCGGTAGGCTCGGCTCGAGCTCGCGCGCCATGAGCGTCCGAGTGCAGATTCCT 72  
Db 15 GACCGTAGGCTCGGCTCGAGCTCGCGCGCCATGAGCGTCCGAGTGCAGATTCCT 74  
Qy 73 CGCCGAGAAGAGCTGTTACCATTTATCCCAACTTCAGTCTGACAAAGATCTACCTCAT 132  
Db 75 CGCCGAGAAGAGCTGTTACCATTTATCCCAACTTCAGTCTGACAAAGATCTACCTCAT 134  
Qy 133 CGGGGGGAGCTGGGGCTTTTAACTTACCTGTTTACCGTGGAGTGCCTCGGCGGC 192  
Db 135 CGGGGGGAGCTGGGGCTTTTAACTTACCTGTTTACCGTGGAGTGCCTCGGCGGC 194  
Qy 193 GATTAACTTGAACAAAGACAGAAATGTCGCTGCTCCCTCCAGTGGATGTAGTA 252  
Db 195 GATTAACTTGAACAAAGACAGAAATGTCGCTGCTCCCTCCAGTGGATGTAGTA 254  
Qy 253 AAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAAGAACTTTTAC 300  
Db 255 AAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAAGAACTTTTAC 302

# RESULT 15 BE251621 LOCUS

DEFINITION 601110640F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:3351416 5',  
mRNA sequence.

ACCESSION BE251621  
VERSION BE251621.1 GI:9121737  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 729)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC  
DNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (L1NL)  
DNA Sequencing by: MGC Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/L1NL at: [Image.l1nl.gov](http://Image.l1nl.gov)  
Plate: L1NCM148 row: p column: 09  
High quality sequence spot: 628

## FEATURES

BASE COUNT	ORIGIN	Query Match	Score	DB	Length	729;
181 a	193 c	214 g	141 t			
		Best Local Similarity	99.3%;	Pred. No. 2e-66;		
		Matches 286;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	13	GTCCGTGAGCCTCTGGCCCTGCACGCTCGCGCGCCCATGGACGCTGCCAGATCGCAATTCCT	72			
Db	8	GACCGTGAGCCTCTGGCCCTGCACGCTCGCGCGCCCATGGACGCTGCCAGATCGCAATTCCT	67			
QY	73	CGCCGAGAAGGAGCTGGTTACCACTATTCCCAACTTCAGTCTGGACAAGATCTACCTCAT	132			
Db	68	CGCCGAGAAGGAGCTGGTTACCACTATTCCCAACTTCAGTCTGGACAAGATCTACCTCAT	127			
QY	133	CGGGGGGACCTGGGGCCCTTTTAACCGCTGGTTTACCCGTGGAAGTGCCTGTGGCTGGC	192			
Db	128	CGGGGGGACCTGGGGCCCTTTTAACCGCTGGTTTACCCGTGGAAGTGCCTGTGGCTGGC	187			
QY	193	GATTAACCTGAAACAAGACAGAAATGTCCCTGTCTCCCTCCAGATGGATGTGAGA	252			
Db	188	GATTAACCTGAAACAAGACAGAAATGTCCCTGTCTCCCTCCAGATGGATGTGAGA	247			
QY	253	AAAGTTGGAGAGATGAGGGATCATGAACGAAAGGAAACCTTTTAC	300			
Db	248	AAAGTTGGAGAGATGAGGGATCATGAACGAAAGGAAACCTTTTAC	295			

Search completed: October 14, 2003, 14:09:09  
Job time : 1278.67 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 11:32:16 ; Search time 1272.67 Seconds  
(without alignments)  
5729.185 Million cell updates/sec

Title: US-09-854-124-5

Perfect score: 300

Sequence:

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

al number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba:\*\*

2: em\_esthum:\*\*

3: em\_estlin:\*\*

4: em\_estnu:\*\*

5: em\_estov:\*\*

6: em\_estpl:\*\*

7: em\_estro:\*\*

8: em\_hc:\*\*

9: gb\_est1:\*\*

10: gb\_est2:\*\*

11: gb\_hc:\*\*

12: gb\_est3:\*\*

13: gb\_est4:\*\*

14: gb\_est5:\*\*

15: em\_estfun:\*\*

16: em\_estom:\*\*

17: em\_gss\_hum:\*\*

18: em\_gss\_inv:\*\*

19: em\_gss\_pin:\*\*

20: em\_gss\_vrt:\*\*

21: em\_gss\_fun:\*\*

22: em\_gss\_mam:\*\*

23: em\_gss\_mus:\*\*

24: em\_gss\_pro:\*\*

25: em\_gss\_rod:\*\*

26: em\_gss\_phg:\*\*

27: em\_gss\_vrl:\*\*

28: gb\_gss1:\*\*

29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	300	100.0	533	14	CB112523 K-EST0154
C 2	300	100.0	559	13	BM126492 if06h06.y
C 3	300	100.0	659	12	BU729618 UT-E-CQ1-
C 4	300	100.0	716	10	BE382866 601297755

5	300	100.0	801	12	BI196248
6	300	100.0	876	10	BI181130
7	300	100.0	896	10	BE562088
8	300	100.0	914	10	BE547290
9	300	100.0	943	10	BF683514
10	300	100.0	970	9	AL560669
11	300	100.0	1007	12	BM449472
12	300	100.0	1064	13	BQ277667
13	300	100.0	1071	13	BU184963
14	300	100.0	1071	9	AL560880
15	300	100.0	1201	9	AL582250
16	299.6	99.9	1201	9	AL582217
17	299	99.7	1810	12	BQ009838
18	298.8	99.6	1193	9	AL529785
19	298.4	99.5	488	14	CB160336
20	298.4	99.5	696	9	AW249012
21	298.4	99.5	712	13	BU625683
22	298.4	99.5	714	10	BG104289
23	298.4	99.5	716	12	BG825252
24	298.4	99.5	730	10	BE296429
25	298.4	99.5	790	10	BE795306
26	298.4	99.5	866	10	BF795157
27	298.4	99.5	867	13	BQ233393
28	298.4	99.5	891	11	BC022839
29	298.4	99.5	921	10	BE799212
30	298.4	99.5	932	10	BE796384
31	298.4	99.5	1028	10	BE561044
32	298.4	99.5	1030	10	BF311745
33	298.4	99.5	1182	11	AF125098
34	298	99.3	833	13	BU597296
35	297.4	99.1	494	14	H64860
36	296.8	98.9	895	13	BU539659
37	293.2	97.7	1056	13	BU456910
38	292.4	97.5	667	9	AU126087
39	292	97.3	966	9	AL582077
40	291.4	97.1	1201	9	AL562756
41	290	96.7	649	10	BE514071
42	289.6	96.5	1007	10	BE795838
43	288.6	96.2	1201	9	AL526847
44	288	96.0	761	12	BI257993
45	287.4	95.8	966	10	BE791539

ALIGNMENTS

RESULT 1  
CB112523/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CB112523  
K-EST0154352 L6ChoCK0 Homo sapiens CDNA clone L6ChoCK0-4-D04 5',  
mRNA sequence.  
CB112523  
CB112523.1 GI:27938330  
EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo;  
1 (bases 1 to 533)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
21C Frontier Korean EST Project 2001  
Unpublished  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 4 row: D column: 04  
High quality sequence stop: 533.

RESULT 3  
BU729618/c



```

BASE COUNT      149 a   159 c   151 g   200 t
ORIGIN
      iag_seq=CCATTAAGTG
Query Match      100.0%; Score 300; DB 13; Length 659;
Best Local Similarity 100.0%; Pred. No. 1.9e-72;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY .      1  ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60
          |||||
Db      509  ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 450
          |||||
QY      61  CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTCACCTTGA 120

```

unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-k@mail.nih.gov  
 Tissue Procurement: AFCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM311 row: n column: 13  
 High quality sequence stop: 651.

```

FEATURES
source
      1. 7716
      Location/Qualifiers
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone_="IMAGE:3627852"
        /tissue_type="neuroblastoma"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_19"
        /note="Organ: brain; Vector: pOT87; Site_1: XhoI; Site_2:
        EcoRI; cDNA made by oligo-dT priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGCACGAG(G). Library constructed by Ling Hong
        in the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies).
        Note: this is a NIH_MGC Library."
      185 a 179 c 195 t 157 t
BASE COUNT

```

```

-BASE COUNT      185 a      179 c      195 g      157 t
-ORIGIN

Query Match      100.0%; Score 300; DB 10; Length 716;
Best Local Similarity 100.0%; Pred. No. 2e-72;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACGAAATCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCGAGTGT 60
        |||||
Db      337  ACGAAATCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCGGACTGT 396
        |||||

QY      61  CTGCTCACAGCTTTGTGAGACACAGAGGCACATGCCAAGCTGGATACTTGACCTTGA 120
        |||||
Db      397  CTGCTCACAGCTTTGTGAGACACAGAGGCACATGCCAAGCTGGATACTTGACCTTGA 456
        |||||

```

121 TGGAGATCAACACCGGGGACTTTCTCCACACAGCGCTCAACACATGTACAACCTCC 180  
 457 TGGAGATCAACACCGGGGACTTTCTCCACACAGCGCTCAACACATGTACAACCTCC 516  
 181 GCACGAACCTCCAGGCTCTGGAGAGTACTCAGTCTCAGGACTCTAGAGAAAGGCTGTGT 240  
 517 GCACGAACCTCCAGGCTCTGGAGAGTACTCAGTCTCAGGACTCTAGAGAAAGGCTGTGT 576  
 241 GCAGCGGCTTCTCTGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTTGG 300  
 577 GCAGCGGCTTCTCTGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTTGG 636

RESULT 5  
 BI196248  
 LOCUS  
 DEFINITION 602754709F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4890147 5',  
 mRNA sequence.  
 ACCESSION BI196248  
 VERSION  
 WORDS  
 EST. BI196248.1 GI:14651268  
 ORGANCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NIH-MGC http://mgc.nci.nih.gov/  
 1 (bases 1 to 801)  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: L1CMI767 row: b column: 04  
 High quality sequence stop: 781.

FEATURES  
 source  
 1..801  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4890147"  
 /tissue\_type="neuroblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_19"  
 /note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZIP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

BASE COUNT 208 a 196 c 216 g 181 t  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 12; Length 801;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 DB 356 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 415  
 QY 61 CTGCTCACAGCTTTGTGAGACACGAGGACATGCCAAGCTGGATATACTTGACCTTGA 120  
 DB 416 CTGCTCACAGCTTTGTGAGACACGAGGACATGCCAAGCTGGATATACTTGACCTTGA 475  
 QY 121 TGGAGATCAACACCGGGGACTTTCTCCACACAGCGCTCAACACATGTACAACCTCC 180

Query Match 100.0%; Score 300; DB 12; Length 801;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 DB 356 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 415  
 QY 61 CTGCTCACAGCTTTGTGAGACACGAGGACATGCCAAGCTGGATATACTTGACCTTGA 120  
 DB 416 CTGCTCACAGCTTTGTGAGACACGAGGACATGCCAAGCTGGATATACTTGACCTTGA 475  
 QY 121 TGGAGATCAACACCGGGGACTTTCTCCACACAGCGCTCAACACATGTACAACCTCC 180

Db 476 TGGAGATCAACACCGGGGACTTTCTCCACACAGCGCTCAACACATGTACAACCTCC 535  
 QY 181 GCACGAACCTCCAGGCTCTGGAGAGTACTCAGTCTCAGGACTCTAGAGAAAGGCTGTGT 240  
 Db 536 GCACGAACCTCCAGGCTCTGGAGAGTACTCAGTCTCAGGACTCTAGAGAAAGGCTGTGT 595  
 QY 241 GCAGCGGCTTCTCTGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTTGG 300  
 Db 596 GCAGCGGCTTCTCTGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTTGG 655

RESULT 6  
 BG181130  
 LOCUS  
 DEFINITION 602329264F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4430823 5',  
 mRNA sequence.  
 ACCESSION BG181130  
 VERSION  
 WORDS  
 EST. BG181130.1 GI:12687833  
 ORGANCE Homo sapiens (human)  
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NIH-MGC http://mgc.nci.nih.gov/  
 1 (bases 1 to 876)  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM10184 row: o column: 16  
 High quality sequence stop: 674.

FEATURES  
 source  
 1..876  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4430823"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_91"  
 /note="Organ: prostate; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.4 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT 246 a 217 c 214 g 199 t  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 10; Length 876;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 DB 115 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 174  
 QY 61 CTGCTCACAGCTTTGTGAGACACGAGGACATGCCAAGCTGGATATACTTGACCTTGA 120  
 DB 175 CTGCTCACAGCTTTGTGAGACACGAGGACATGCCAAGCTGGATATACTTGACCTTGA 234  
 QY 121 TGGAGATCAACACCGGGGACTTTCTCCACACAGCGCTCAACACATGTACAACCTCC 180  
 DB 235 TGGAGATCAACACCGGGGACTTTCTCCACACAGCGCTCAACACATGTACAACCTCC 294  
 QY 181 GCACGAACCTCCAGGCTCTGGAGAGTACTCAGTCTCAGGACTCTTAGAGAAAGGCTGTGT 240

Db 295 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 354  
 QY 241 GCAGGCGGCTTGTGGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 300  
 Db 355 GCAGGCGGCTTGTGGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 414

RESULT 7  
 BE562088 896 bp mRNA linear EST 15-AUG-2000  
 LOCUS 601345038F1 NIH\_MCC\_8 Homo sapiens cDNA clone IMAGE:3678080 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE562088  
 VERSION BE562088.1 GI:9805808  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 896)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LCM354 row: k column: 09  
 High quality sequence start: 24  
 High quality sequence stop: 840.

FEATURES  
 source  
 1..896  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3678080"  
 /tissue\_type="Burkitt lymphoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_8"  
 /note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 218 a 230 c 251 g 197 t

Query Match 100.0%; Score 300; DB 10; Length 896;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAAATCCGGACCTTGTGACAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 Db 438 ACGAAATCCGGACCTTGTGACAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 497  
 QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACTTGACCTTGA 120  
 Db 498 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACTTGACCTTGA 557  
 QY 121 TGGAGATCAACACACGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 180  
 Db 558 TGGAGATCAACACACGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 617  
 QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240

Db 618 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 677  
 QY 241 GCAGGCGGCTTGTGGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 300  
 Db 678 GCAGGCGGCTTGTGGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 737

RESULT 8  
 BE547290 914 bp mRNA linear EST 09-AUG-2000  
 LOCUS 601073715F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:3460010 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE547290  
 VERSION BE547290.1 GI:9775935  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 914)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAMB453 row: m column: 03  
 High quality sequence stop: 658.

FEATURES  
 source  
 1..914  
 Location/Qualifiers

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3460010"  
 /tissue\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_12"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site:1: NotI;  
 Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."  
 254 a 227 c 238 g 195 t

BASE COUNT 254 a 227 c 238 g 195 t  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 10; Length 914;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAAATCCGGACCTTGTGACAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 Db 104 ACGAAATCCGGACCTTGTGACAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 163  
 QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACTTGACCTTGA 120  
 Db 164 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACTTGACCTTGA 223  
 QY 121 TGGAGATCAACACACGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 180  
 Db 224 TGGAGATCAACACACGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 283  
 QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
 Db 284 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 343  
 QY 241 GCAGGCGGCTTGTGGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 300  
 Db 344 GCAGGCGGCTTGTGGGGGATGTAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 403

*Cervical  
 cell line*

RESULT 9  
BF683514  
LOCUS  
DEFINITION  
602139714F1 NIH\_MGC\_46 Homo sapiens cDNA clone IMAGE:4300947 5',  
mRNA sequence.  
ACCESSION  
BF683514  
VERSION  
BF683514.1 GI:11968922  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 943)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
JOURNAL  
National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L10M1160 row: 1 column: 04  
High quality sequence stop: 709.  
Location/Qualifiers  
1. 943  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4300947"  
/tissue\_type="leiomyosarcoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_46"  
/note="Organ: uterus; Vector: pOTB7; Site:1; XhoI; Site:2;  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 214 a 253 c 272 g 204 t  
ORIGIN

Query Match 100.0%; Score 300; DB 10; Length 943;  
Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
DB 369 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 428

QY 61 CTGCTCACAGCTTTGTGAGACAGCAGGACACATGCCAAGCTGGATTAAGTTCACCTTGA 120  
DB 429 CTGCTCACAGCTTTGTGAGACAGCAGGACACATGCCAAGCTGGATTAAGTTCACCTTGA 488

QY 121 TGGAGATCAACACACAGCGGAGCTTTCTCACACAAGCGCTCAACACATGTACAACTCC 180  
DB 489 TGGAGATCAACACACAGCGGAGCTTTCTCACACAAGCGCTCAACACATGTACAACTCC 548

QY 181 GCACGAACCTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGTG 240  
DB 549 GCACGAACCTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGTG 608

QY 241 GCAGCGGCTTCTCGGGGATGTGAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTTGG 300  
DB 609 GCAGCGGCTTCTCGGGGATGTGAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTTGG 668

RESULT 10  
AL560669  
LOCUS  
DEFINITION  
AL560669 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
Homo sapiens cDNA clone CS0DL003YI21 5-PRIME, mRNA sequence.  
ACCESSION  
AL560669  
VERSION  
AL560669.2 GI:31284799  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 970)  
AUTHORS  
Li, W.B., Gruber, C., Jessee, J., and Pollayes, D.  
JOURNAL  
Full-length cDNA libraries and normalization  
COMMENT  
Unpublished  
On Feb 15, 2001 this sequence version replaced gi:12907354.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7987.r For  
more information about this cluster, see  
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DL003AE11Q1&cluster=7987.r)  
cgi-bin/cluster.cgi?seq=CS0DL003AE11Q1&cluster=7987.r. Contact :  
Feng Liang Email : [liang@lifetech.com](mailto:liang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>  
Faraday Avenue Genoscope sequence ID : CS0DL003AE11Q1.  
Location/Qualifiers  
1. 970  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL003YI21"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 257 a 243 c 251 g 219 t  
ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 970;  
Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
DB 419 ACAGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 478

QY 61 CTGCTCACAGCTTTGTGAGACAGCAGGACACATGCCAAGCTGGATTAAGTTCACCTTGA 120  
DB 479 CTGCTCACAGCTTTGTGAGACAGCAGGACACATGCCAAGCTGGATTAAGTTCACCTTGA 538

QY 121 TGGAGATCAACACACAGCGGAGCTTTCTCACACAAGCGCTCAACACATGTACAACTCC 180  
DB 539 TGGAGATCAACACACAGCGGAGCTTTCTCACACAAGCGCTCAACACATGTACAACTCC 598

QY 181 GCACGAACCTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGTG 240  
DB 599 GCACGAACCTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGTG 658

QY 241 GCAGCGGCTTCTCGGGGATGTGAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTTGG 300  
DB 659 GCAGCGGCTTCTCGGGGATGTGAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTTGG 718

```

RESULT 11
BM449472
LOCUS
DEFINITION
AGENCOURT_6400913 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493679
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM12117 row: m column: 08
High quality sequence stop: 711.
Location/Qualifiers
1..1007
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5493679"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/Note="Organ: eye; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 259 a 254 c 262 g 231 t 1 others
ORIGIN
Query Match 100.0%; Score 300; DB 12; Length 1007;
Best Local Similarity 100.0%; Pred. No. 2.4e-72;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACGAAATCCGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60
357 ACGAAATCCGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 416
QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 120
417 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 476
QY 121 TGGAGATCAACACACCGCGGACTTTCCTCACACAGCGCTCAACACATGTACAAACTCC 180
477 TGGAGATCAACACACCGCGGACTTTCCTCACACAGCGCTCAACACATGTACAAACTCC 536
QY 181 GCACAACTCCAGCCTCTGGAGAGTACTAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 240
537 GCACAACTCCAGCCTCTGGAGAGTACTAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 596
QY 241 GCAGCGGCTTCCTGGGGGATGTAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTGG 300
597 GCAGCGGCTTCCTGGGGGATGTAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTGG 656

RESULT 12
BQ277667
LOCUS
DEFINITION
AGENCOURT_7048107 NIH_MGC_109 Homo sapiens cDNA clone IMAGE:5804514
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ruben Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2042 row: d column: 19
High quality sequence stop: 675.
Location/Qualifiers
1..1064
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5804514"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/Note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 275 a 268 c 273 g 245 t 3 others
ORIGIN
Query Match 100.0%; Score 300; DB 13; Length 1064;
Best Local Similarity 100.0%; Pred. No. 2.5e-72;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACGAAATCCGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60
364 ACGAAATCCGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 423
QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 120
424 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 483
QY 121 TGGAGATCAACACACCGCGGACTTTCCTCACACAGCGCTCAACACATGTACAAACTCC 180
484 TGGAGATCAACACACCGCGGACTTTCCTCACACAGCGCTCAACACATGTACAAACTCC 543
QY 181 GCACAACTCCAGCCTCTGGAGAGTACTAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 240
544 GCACAACTCCAGCCTCTGGAGAGTACTAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 603
QY 241 GCAGCGGCTTCCTGGGGGATGTAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTGG 300
604 GCAGCGGCTTCCTGGGGGATGTAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTGG 663

RESULT 13
BU184963
LOCUS
DEFINITION
AGENCOURT_7186943 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6002100
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ruben Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2042 row: d column: 19
High quality sequence stop: 675.
Location/Qualifiers
1..1064
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5804514"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/Note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 275 a 268 c 273 g 245 t 3 others
ORIGIN
Query Match 100.0%; Score 300; DB 13; Length 1064;
Best Local Similarity 100.0%; Pred. No. 2.5e-72;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACGAAATCCGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60
364 ACGAAATCCGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 423
QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 120
424 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 483
QY 121 TGGAGATCAACACACCGCGGACTTTCCTCACACAGCGCTCAACACATGTACAAACTCC 180
484 TGGAGATCAACACACCGCGGACTTTCCTCACACAGCGCTCAACACATGTACAAACTCC 543
QY 181 GCACAACTCCAGCCTCTGGAGAGTACTAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 240
544 GCACAACTCCAGCCTCTGGAGAGTACTAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 603
QY 241 GCAGCGGCTTCCTGGGGGATGTAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTGG 300
604 GCAGCGGCTTCCTGGGGGATGTAGCGCTCAGGAGTGTAGGACTCTCGTGGTCTGG 663

```

VERSION BUI84963.1 GI:22698947  
 .KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1071)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LCM2269 row: m column: 13  
 High quality sequence stop: 718.  
 Location/Qualifiers  
 1. 1071  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6002100"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_100"  
 /note="Organ: liver; Vector: pORF7; Site\_1: XhoI; Site\_2:  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH-MGC  
 Library."

URES  
 source

BASE COUNT 283 a 262 c 275 g 246 t 5 others  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 13; Length 1071;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60  
 Db 359 ACGAATCCGGACCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 418  
 61 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 120  
 Db 419 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 478  
 121 TGGAGATCAACACACAGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACCTCC 180  
 Db 479 TGGAGATCAACACACAGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACCTCC 538  
 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGAGCTTCTAGAGAAAGGCTGTGT 240  
 Db 539 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGAGCTTCTAGAGAAAGGCTGTGT 598  
 241 GCAGCGGCTTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 300  
 Db 599 GCAGCGGCTTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 658

RESULT 14  
 AL560880  
 LOCUS 1201 bp mRNA linear EST 31-MAY-2003  
 DEFINITION AL560880 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 ACCESION Homo sapiens cDNA clone CSODL005YK19 5-PRIME, mRNA sequence.  
 AL560880

AL560880.2 GI:31285009  
 EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1201)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 JOURNAL  
 COMMENT  
 On Feb 15, 2001 this sequence version replaced gi:12907768.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 7987.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSODL005AF100P1&cluster=7987.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSODL005AF100P1.  
 Location/Qualifiers  
 1. 1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODL005YK19"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 BASE COUNT 322 a 267 c 292 g 283 t 37 others  
 ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 1201;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60  
 Db 426 ACGAATCCGGACCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 485  
 61 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 120  
 Db 486 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGA 545  
 121 TGGAGATCAACACACAGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACCTCC 180  
 Db 546 TGGAGATCAACACACAGCGGACTTTCCTCACACAAGCGCTCAACACATGTACAACCTCC 605  
 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGAGCTTCTAGAGAAAGGCTGTGT 240  
 Db 606 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGAGCTTCTAGAGAAAGGCTGTGT 665  
 241 GCAGCGGCTTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 300  
 Db 666 GCAGCGGCTTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 725

RESULT 15  
 AL582250/c  
 LOCUS 1201 bp mRNA linear EST 01-JUN-2003  
 DEFINITION AL582250 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 ACCESION Homo sapiens cDNA clone CSODL005YK11 3-PRIME, mRNA sequence.  
 AL582250  
 VERSION AL582250.2 GI:31320467  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 16, 2001 this sequence version replaced gi:12950048.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7987.r for  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DL005CC06NP1&cluster=7987.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DL005CC06NP1.  
Location/Qualifiers  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL005YF11"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 291 a 238 c 281 g 293 t 98 others  
ORIGIN

## FEATURES

source

Query Match 100.0%; Score 300; DB 9; Length 1201;  
Best Local Similarity 100.0%; Pred. No. 2.6e-72;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
|||||  
Db 746 ACGAAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 687  
|||||

QY 61 CTGCTGACACTTGTGTAGACAGCAGGAGGCACATGCCAAGCTGGAGTAACCTTGACCTGA 120  
|||||  
686 CTGCTGACACTTGTGTAGACAGCAGGAGGCACATGCCAAGCTGGAGTAACCTTGACCTGA 627  
|||||

121 TGGAGATCAACACACAGCGGACTTTCCTCACACAGGCTCAACCATGTACAAACTCC 180  
|||||

Db 626 TGGAGATCAACACACAGCGGACTTTCCTCACACAGGCTCAACCATGTACAAACTCC 567  
|||||

QY 181 GCACGAACCTCCAGCCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
|||||

Db 566 GCACGAACCTCCAGCCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 507  
|||||

QY 241 GCAGGGCGCTTGTGGGGGATGTAGCGGCTCAGGACGTATGAGGTACTCGTGGTCTGG 300  
|||||

Db 506 GCAGGGCGCTTGTGGGGGATGTAGCGGCTCAGGACGTATGAGGTACTCGTGGTCTGG 447  
|||||

Search completed: October 14, 2003, 14:09:03  
Job time : 1279.67 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 12:04:45 ; Search time 157.333 Seconds  
(without alignments)  
5147.234 Million cell updates/sec

Title: US-09-854-124-5  
Perfect score: 300  
Sequence: 1 acgaatccgaccctggtc.....tgagtgactcgtgtttctgg 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq 15Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	300	100.0	300	ABQ78750	Differentially exp
2	300	100.0	858	AAH14111	Human cDNA sequenc
3	300	100.0	1171	ACAS4679	Human NF-kappaB as
4	300	100.0	1210	RAC77820	Human cancer assoc
5	292.4	97.5	667	AAH05624	Human cDNA clone (
6	252.8	84.3	1190	AAH87106	DNA encoding novel
7	252.8	84.3	1499	AAH87104	DNA encoding novel
8	240	80.0	678	AAH87103	DNA encoding novel

9	240	80.0	678	23	AAH87105	DNA encoding novel
10	214.4	71.5	300	24	ABQ78752	Differentially exp
11	140	46.7	454	24	ABL68902	Kidney cancer rela
12	100.6	33.5	603	24	ABQ59992	Human colon cancer
13	77.6	25.9	1257	23	AAH83051	DNA encoding novel
c 14	66	22.0	362	24	ABV88973	Human colon cancer
c 15	66	22.0	362	24	ABK45131	CDNA encoding colo
16	58.8	19.6	649	24	ABQ59972	Human colon cancer
17	53.8	17.9	612	23	ABL19607	Drosophila melanog
18	53.8	17.9	2670	23	ABL19606	Drosophila melanog
19	53.8	17.9	4504	23	ABL05636	Drosophila melanog
c 20	49.4	16.5	51	22	AAH27661	Human SNP oligonuc
21	39.6	13.2	734	22	AAH08244	Human cDNA clone (
22	39.6	13.2	940	23	AAH72563	DNA encoding novel
23	39.6	13.2	1143	25	ABX71197	Novel human cDNA s
24	39.6	13.2	3071	22	AAH16628	Human cDNA sequenc
25	36	12.0	654	24	ABN61061	Human cancer relat
26	35	11.7	2241	19	AAV68070	Maize-optimised DN
27	35	11.7	2370	19	AAV68069	Maize-optimised DN
28	35	11.7	2403	17	AAH13955	Maize-optimised VI
29	35	11.7	2403	18	AAH74004	Maize-optimised-B.
30	35	11.7	2403	19	AAV68065	Maize-optimised DN
31	35	11.7	2403	17	AAH16177	Maize optimised DN
32	35	11.7	2444	17	AAH13964	Maize optimised DN
33	35	11.7	2444	18	AAH74011	Maize optimised DN
34	35	11.7	2444	19	AAV16190	B. cereus VIP3A(a)
35	33.8	11.3	360	21	AAZ50397	DNA for maize opti
36	33.2	11.1	1206	24	ABN81296	Human interferon-1
37	33.2	11.1	1363	24	ABN81295	Ancyclobacter form
38	32.8	10.9	468	22	ABA42112	Human breast cell
39	32.8	10.9	468	22	ABA52534	Human foetal liver
40	32.8	10.9	468	22	ABA22323	Probe #789 for gen
41	32.8	10.9	468	22	AAK00797	Human brain expres
42	32.8	10.9	468	22	AAK26250	Human bone marrow
43	32.8	10.9	468	22	AAH10883	Probe #816 for gen
44	32.8	10.9	468	22	AAH32143	Probe #829 used to
45	32.8	10.9	468	22	AAH00806	Probe #797 used to

ALIGNMENTS

RESULT 1  
ABQ78750  
ID ABQ78750 standard; cDNA; 300 BP.  
XX AC ABQ78750;  
XX DT 05-DEC-2002 (first entry)  
XX DE Differentially expressed gene RTA00000684F.e.07.1.  
XX KW Differentially expressed gene; cancer; breast cancer; colon cancer;  
KW lung cancer; prostate cancer; expressed sequence tag; EST;  
KW heat shock factor binding protein; CGI-122 gene; quiescin Q6;  
KW moderately-differentiated endometrial adenocarcinoma; TGIF protein;  
KW treacher collins syndrome protein; annexin IV; cyclophilin C;  
KW MHC class I lymphocyte antigen; HLA-E class I mRNA;  
KW glomerulosclerosis gene; 265 proteasome subunit p55;  
KW gamma interferon inducible protein; cyclin-dependent protein kinase;  
KW mitochondrial dodecenoyl-CoA delta-isomerase;  
KW serine hydroxymethyltransferase;  
KW DNA-damage-inducible RNA binding protein; ss.  
OS Homo sapiens.  
XX US2002076735-A1.  
XX 20-JUN-2002.  
XX PF 10-MAY-2001; 2001US-0854124.  
XX PR 25-SEP-1998; 98US-101900P.



CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX  
 SQ Sequence 858 BP; 227 A; 214 C; 222 G; 195 T; 0 other;

Query Match 100.0%; Score 300; DB 22; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-86;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAATCCGGACCTGCTCAAGGATATGTGGGACACCTCGTATAGCCAACTCCGAGTGT 60  
 Db |||||

368 ACGAATCCGGACCTGCTCAAGGATATGTGGGACACCTCGTATAGCCAACTCCGAGTGT 427

QY 61 CTGCTGACAGCTTTCTGACACAGCAGGAGGCACATGCCAAGCTGGATACCTTGACCTTCA 120

Db |||||

428 CTGCTGACAGCTTTCTGACACAGCAGGAGGCACATGCCAAGCTGGATACCTTGACCTTCA 487

QY 121 TGGAGATCAACACACCGCGGACTTCTCTCACAAAGCGCTCAACACCATGTACAAACTCC 180

Db |||||

488 TGGAGATCAACACACCGCGGACTTCTCTCACAAAGCGCTCAACACCATGTACAAACTCC 547

QY 181 GCACCACTCTCAGCTCTGGAGAGTACTCAGTCTCAGGACCTCTAGAGAAAGCGCTGT 240

Db |||||

548 GCACCACTCTCAGCTCTGGAGAGTACTCAGTCTCAGGACCTCTAGAGAAAGCGCTGT 607

QY 241 GCAGCGGGTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTGG 300

Db |||||

608 GCAGCGGGTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTGG 667

RESULT 3  
 ID ACA54679 standard; cDNA; 1171 BP.

XX ACA54679;

AC ACA54679;

XX 05-JUN-2003 (first entry)

XX Human NF-kappaB associated polynucleotide sequence #119.

XX Human; nuclear factor-kappaB; NF-kappaB; immune disorder; cancer;  
 KW inflammatory disorder; apoptosis; hepatic disorder; Hodgkin's lymphoma;  
 KW haematopoietic tumour; hyper-IgM syndrome; viral infection; asthma;  
 KW hypohidrotic ectodermal dysplasia; human immunodeficiency virus; HIV;  
 KW X-linked anhidrotic ectodermal dysplasia; al incontinentia pigmenti;  
 KW influenza; rheumatoid arthritis; inflammatory bowel disease; colitis;  
 KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;  
 KW experimental allergic encephalomyelitis; autoimmune disorder; wound;  
 KW hyper immune activity; acute phase response; hypercongenital condition;  
 KW birth defect; necrotic lesion; organ transplant rejection; pancreas;  
 KW signal transduction; hyperproliferative disorder; diabetes mellitus;  
 KW vitamin B12 malabsorption; neurological disorder; Huntington's chorea;  
 KW Turner's syndrome; bacterial infection; cardiovascular disorder;  
 KW infertility; psoriasis; haemolytic anaemia; antiinflammatory; anti-HIV;  
 KW cytostatic; hepatotropic; virucide; antirheumatic; antiarthritic;  
 KW antiasthmatic; immunomodulator; antidiabetic; antiallergic;  
 KW neuroprotective; immunosuppressive; vulnerable; antibacterial;  
 KW antifertility; antianemic; antipsoriatic; cerebroprotective;  
 KW cardiant; antiarteriosclerotic; gene; ss.

XX Homo sapiens.

OS WO200286076-A2.

XX 31-OCT-2002.

XX 19-APR-2002; 2002WO-US12636.

XX 19-APR-2001; 2001US-284962P.

XX 26-APR-2001; 2001US-286645P.

PR

PR 09-JAN-2002; 2002US-346986P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Carman J, Feder J, Nadler S;

XX WPI; 2003-093119/08.

DR P-PSDB; ABU69618.

XX Novel NF-kappaB-associated polypeptides and polynucleotides useful for

XX diagnosing, treating and preventing cancer, hepatic disorders, aberrant

XX apoptosis, viral infections, autoimmune disorders, asthma and stroke -

XX Claim 4; Fig 16; 608pp; English.

XX The present invention relates to the isolation of human nuclear

XX factor-kappaB (NF-kappaB) associated polypeptides and polynucleotides.

XX The NF-kappaB associated polypeptide and polynucleotide sequences

XX are useful for preventing, treating or ameliorating various disorders

XX including immune disorders, inflammatory disorders, cancers,

XX disorders relating to aberrant apoptosis, hepatic disorders,

XX Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM syndromes,

XX hypohidrotic ectodermal dysplasia, X-linked anhidrotic ectodermal

XX dysplasia, immunodeficiency, al incontinentia pigmenti, viral

XX infections (e.g. those caused by human immunodeficiency virus (HIV),

XX human T-cell lymphotropic virus (HTLV), hepatitis B, hepatitis C,

XX inflammatory bowel disease, colitis, asthma, atherosclerosis, cachexia,

XX euthyroid sick syndrome, stroke, experimental allergic encephalomyelitis,

XX (EAE), autoimmune disorders, disorders related to hyper immune activity,

XX disorders related to aberrant acute phase responses, hypercongenital

XX conditions, birth defects, necrotic lesions, wounds, organ transplant

XX rejection, disorders related to aberrant signal transduction,

XX hyperproliferative disorders, diseases of the pancreas (e.g. diabetes

XX mellitus, vitamin B12 malabsorption), neurological disorders (e.g.

XX Huntington's chorea), Turner's syndrome, bacterial infections,

XX cardiovascular disorders, infertility, psoriasis and haemolytic anaemia.

XX The present sequence represents a human NF-kappaB associated

XX polynucleotide of the invention.

XX Sequence 1171 BP; 296 A; 298 C; 309 G; 268 T; 0 other;

QY Query Match 100.0%; Score 300; DB 25; Length 1171;

Db Best Local Similarity 100.0%; Pred. No. 1.1e-85;

QY Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACGAATCCGGACCTGCTCAAGGATATGTGGGACACCTCGTATAGCCAACTCCGAGTGT 60

Db |||||

408 ACGAATCCGGACCTGCTCAAGGATATGTGGGACACCTCGTATAGCCAACTCCGAGTGT 467

QY 61 CTGCTGACAGCTTTCTGACACAGCAGGAGGCACATGCCAAGCTGGATACCTTGACCTTGA 120

Db |||||

468 CTGCTGACAGCTTTCTGACACAGCAGGAGGCACATGCCAAGCTGGATACCTTGACCTTGA 527

QY 121 TGGAGATCAACACACCGCGGACTTCTCTCACAAAGCGCTCAACACCATGTACAAACTCC 180

Db |||||

528 TGGAGATCAACACACCGCGGACTTCTCTCACAAAGCGCTCAACACCATGTACAAACTCC 587

QY 181 GCACCACTCTCAGCTCTGGAGAGTACTCAGTCTCAGGACCTCTAGAGAAAGCGCTGT 240

Db |||||

588 GCACCACTCTCAGCTCTGGAGAGTACTCAGTCTCAGGACCTCTAGAGAAAGCGCTGT 647

QY 241 GCAGCGGGTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTGG 300

Db |||||

648 GCAGCGGGTCTCTGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTGG 707

RESULT 4

AACT7820

ID AAC77820 standard; cDNA; 1210 BP.

XX AAC77820;

AC AAC77820;

XX

08-FEB-2001 (first entry)  
Human cancer associated gene sequence SEQ ID NO:214.  
Human; cancer associated gene; cancer antigen; detection; cancer;  
diagnosis; cytostatic; proliferative; vulnary; immunomodulator;  
antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;  
antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;  
dermatological; neuroprotective; thrombolytic; coagulant; neotropic;  
vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
immune disorder; haematopoietic cell disorder; autoimmune disorder;  
allergic reaction; graft versus host disease; organ rejection;  
haemostatic; thrombolytic; cardiovascular disorder; infection;  
neurological disease; drug screening; ss.  
Homo sapiens.  
WO200055350-A1.  
21-SEP-2000.  
08-MAR-2000; 2000WO-US05882.  
12-MAR-1999; 99US-0124270.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Ruben SM;  
WPI; 2000-587533/55.  
P-PSDB; AAB43611.  
Novel isolated nucleic acids comprising sequences encoding peptides  
useful for treating or diagnosing e.g. cancer -  
Claim 1; Page 785-786; 2352pp; English.  
AAC77607 to AAC78448 encode the human cancer associated proteins given  
in AAB43398 to AAB44239. The proteins can have activities based on the  
tissues and cells the genes are expressed in. Example of activities  
include: cytostatic; proliferative; vulnary; immunomodulator;  
antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
dermatological; neuroprotective; cardiac; thrombolytic; coagulant;  
neotropic; vasotropic; antipsoriatic and antiangiogenic. The  
polynucleotides and polypeptides can be used for preventing, treating or  
ameliorating medical conditions and diagnosing pathological conditions.  
Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
the present invention may be used to treat immune disorders by activating  
or inhibiting the proliferation, differentiation or mobilisation of  
immune cells, to treat disorders of haematopoietic cells, autoimmune  
disorders, allergic reactions, graft versus host disease and organ  
rejection, modulate haemostatic or thrombolytic activity, modulate  
inflammation, cancers, cardiovascular disorders, neurological disease and  
bacterial or viral infections. The peptides, nucleotides, antibodies,  
agonists and antagonists may be also be used in drug screens. AAC78449 to  
AAC78457 and AAB44240 represent sequences used in the exemplification of  
the present invention.  
SQ Sequence 1210 BP; 320 A; 300 C; 320 G; 268 T; 2 other;  
Query Match 100.0%; Score 300; DB 21; Length 1210;  
Best Local Similarity 100.0%; Pred. No. 1.1e-85;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ACAGAAATCCGACCTGCTCAAGGATATGTGGGACACTCGTATAGCCAAACTCGGCTGT 60  
DB 412 ACAGAAATCCGACCTGCTCAAGGATATGTGGGACACTCGTATAGCCAAACTCGGCTGT 471  
QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 120  
DB 472 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 531

QY 121 TGGAGATCAACACACAGCGGAGCTTTCTCACAACAAGCGCTCAACACATGTACAAATCC 180  
DB 532 TGGAGATCAACACACAGCGGAGCTTTCTCACAACAAGCGCTCAACACATGTACAAATCC 591  
QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGGT 240  
DB 592 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGGT 651  
QY 241 GCAGCGCGCTTGTCTGGGGGATGTGAGCGCTCAGGAGCTGTAGGAGTACTCGTGGTCTGG 300  
DB 652 GCAGCGCGCTTGTCTGGGGGATGTGAGCGCTCAGGAGCTGTAGGAGTACTCGTGGTCTGG 711

RESULT 5  
AAH05624  
ID AAH05624 standard; cDNA; 667 BP.  
XX  
AC AAH05624;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer) SEQ ID NO:2459.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAR-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 1; SEQ ID 2459; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 667 BP; 171 A; 176 C; 177 G; 137 T; 6 other;

Query Match 97.5%; Score 292.4; DB 22; Length 667;  
Best Local Similarity 97.7%; Pred. No. 2.3e-83;  
Matches 293; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGACACATCGTATAGCCAACTCCGAGTGT 60

DB 367 ACGAATCCGGACCTGGTCAAGGATATGTGGACACATCGTATAGCCAACTCCGAGTGT 426

QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGACATGCCAAGCTGGATTAACCTTGGA 120

DB 427 CTGCTGACAGCTTTGTGAGACAGCAGGAGGACATGCCAAGCTGGATTAACCTTGGA 486

QY 121 TGGAGATCAACACACCGGCGCTTCTCCACACAGGCGCTCAACCATGTACAACTCC 180

DB 487 TGGAGATCAACACACCGGCGCTTCTCCACACAGGCGCTCAACCATGTACAACTCC 546

QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCACTCAGGACTTCTAGAGAAAGCGCTGGT 240

DB 547 GCACGAACCTCCAGCCTCTGGAGAGTACTCACTCAGGACTTCTANANAAGGCGCTGGT 606

QY 241 GCAGCGGCTTCTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 300

DB 607 GCAGCGGCTTCTGGGGATGTGAGCGCTCAGGAGCTGATNAAGTACTCGTGGTCTGG 666

RESULT 6  
AAS87106

ID AAS87106 standard; cdNA; 1190 BP.

XX AC AAS87106;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #22910.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABC22919.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 1; SEQ ID NO 22910; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1190 BP; 298 A; 304 C; 310 G; 278 T; 0 other;

Query Match 84.3%; Score 252.8; DB 23; Length 1190;  
Best Local Similarity 98.0%; Pred. No. 1.4e-70;  
Matches 298; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60

DB 413 ACGAATCCGGACCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 472

QY 61 CTGCTGACAGC-TTGTGTGACAGCAGGAGGACATGCCAAGCT-CGATAACTTGACCTT 118

DB 473 CTGCTGACAGC-TTGTGTGACAGCAGGAGGACATGCCAAGCTCGGTAACCTTGACTT 532

QY 119 -GATGAGATCAACACACCGCGG-GACTTTCCTCACACAGCGCTCAACCATGTACAAA 176

DB 533 CGATGAGATCCACACCGCGGTGACTTTCCTCACACAGCGCTCAACCATGTACAAA 592

QY 177 CTCGCCACCACTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTCTAGAGAAAGGCC 236

DB 593 CTCGCCACCACTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTCTAGAGAAAGGCC 652

QY 237 TGGTGCAGCGCGCTTGTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTT 296

DB 653 TGGTGCAGCGCGCTTGTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTT 712

QY 297 CTGG 300

DB 713 CTGG 716

RESULT 7  
AAS87104

ID AAS87104 standard; cdNA; 1499 BP.

XX AC AAS87104;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #22908.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.



Db	580	GCACAACTCCAGCCTTCGGAGAGTACTCACTCAGGACTTCAGAGAAAGGCGCTGCT	639
QY	241	GCAGCGCGCTTCTCGGGGATGTGCGGCTCAGGACGTGA	280
Db	640	GCAGGACGCTTCTCGGGGATGTGA-CGCTCAGGACGTGA	678
RESULT 9			
AA587105			
ID	AAS87105 standard; cDNA; 678 BP.		
XX	AC	AAS87105;	
XX	XX		
XX	DT	13-FEB-2002 (first entry)	
XX	DE	DNA encoding novel human diagnostic protein #22909.	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX	OS	Homo sapiens.	
XX	PN	WO200175067-A2.	
XX	PD	11-OCT-2001.	
XX	XX		
PF	30-MAR-2001; 2001WO-US08631.		
XX	XX		
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Drmanac RT, Liu C, Tang YT;	
XX	DR		
DR	WPI; 2001-639362/73.		
DR	P-PSDB; ABC22918.		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity		
XX	Claim 1; SEQ ID No 22909; 103pp; English.		
XX	CC	The invention relates to isolated polynucleotide (I) and	
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. AAS64197-AAS94564 represent novel human		
CC	diagnostic coding sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX	Sequence 678 BP; 177 A; 185 C; 206 G; 110 T; 0 other;		
Query Match			
Best Local Similarity 80.0%; Score 240; DB 23; Length 678;			
Matches 268; Conservative 0; Mismatches 10; Indels 2; Gaps 2;			

QY	1	ACGAAATCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT	60
Db	401	ACGAAATCGGACCTGGTCAAGGATATATGGACACTCGTATAGCCAAACTCCGAGTGT	460
QY	61	CTGCTGACAGCTTTGTGAGACAGCAGGAGGACATGCCAAGCTGGATTAACCTTGA	120
Db	461	CTGCTGACAGCTTTGTGAGGCAACAGGAGGACATGCCAAGCCGGATTAACCTTGA	520
QY	121	TGGAGATCAACACACAGCGGACTTTCTCACAACAAGCGCTCAACACATGTACAACTCC	180
Db	521	TGGAGATCAACCCAGCGGACTTTCTCACAACAAGCGCTCAACACATGTACAACT-C	579
QY	181	GCACGAACCTCCAGCCTCTGGAGAGTACTCAGCTCAGGACTTCTAGAGAAAGGCGTGT	240
Db	580	GCACGAACCTCCAGCCTCTGGAGAGTACTCAGCTCAGGACTTCTAGAGAAAGGCGTGT	639
QY	241	GCAGCGGCTTCTGGGGATGTGAGCGCTCAGGACGTGA	280
Db	640	GCAGCGGCTTCTGGGGATGTGA-CGCTCAGGACGTGA	678
RESULT 10			
AAQ78752			
ID	ABQ78752 standard; cDNA; 300 BP.		
XX	AC	ABQ78752;	
XX	DT	05-DEC-2002 (first entry)	
XX	DE	Differentially expressed gene RTA00000596f.d.12.1.	
XX	XX		
KW	Differentially expressed gene; cancer; breast cancer; colon cancer;		
KW	lung cancer; prostate cancer; expressed sequence tag; EST;		
KW	heat shock factor binding protein; CGI-122 gene; quiescin Q6;		
KW	moderately-differentiated endometrial adenocarcinoma; TGIF protein;		
KW	treacher collins syndrome protein; annexin IV; cyclophilin C;		
KW	MHC class I lymphocyte antigen; HLA-E class I MRNA;		
KW	glomerulosclerosis gene; 265 proteasome subunit p55;		
KW	gamma interferon inducible protein; cyclin-dependent protein kinase;		
KW	mitochondrial dodecenoyl-CoA delta-isomerase;		
KW	serine hydroxymethyltransferase;		
KW	DNA-damage-inducible RNA binding protein; ss.		
XX	OS	Homo sapiens.	
XX	PN	US2002076735-A1.	
XX	XX		
PD	20-JUN-2002.		
XX	XX		
PF	10-MAY-2001; 2001US-0854124.		
XX	XX		
PR	25-SEP-1998; 98US-101900P.		
PR	22-SEP-1999; 99US-0400947.		
XX	(WILL/) WILLIAMS L T.		
PA	(ESCO/) ESCOBEDO J.		
PA	(INNI/) INNIS M A.		
PA	(GARC/) GARCIA P D.		
PA	(SUDD/) SUDDUTH-KLINGER J.		
PA	(REIN/) REINHARD C.		
PA	(GIES/) GIESE K.		
PA	(RAND/) RANDAZZO F.		
PA	(KENN/) KENNEDY G C.		
PA	(POTD/) POT D.		
PA	(KASS/) KASSAM A.		
PA	(LAWM/) LAWSON G.		
PA	(DRMA/) DRMANAC R.		
PA	(CRKV/) CRKVENJAKOV R.		
PA	(DICK/) DICKSON M.		
PA	(DRMA/) DRMANAC S.		
PA	(LABA/) LABAT I.		
PA	(LESH/) LESHKOWITZ D.		
PA	(KITA/) KITA D.		







XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in AB161664  
CC to AB170110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 454 BP; 128 A; 113 C; 116 G; 97 T; 0 other;

Query Match 46.7%; Score 140; DB 24; Length 454;  
Best Local Similarity 99.3%; Pred. No. 1.2e-34;  
Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 ACGAATCCGGACCTTGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60  
DB 303 ACGAATCCGGACCTTGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362  
QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 120  
DB 363 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 421  
QY 121 TGGAGATCAACACACCGCGGACTTTCTCCACCA 152  
DB 422 TGGAGATCAACACACCGCGGACTTTCTCCACCA 453

RESULT 12  
ABQ59992  
ID ABQ59992 standard; cdna; 603 BP.  
XX  
AC ABQ59992;  
XX  
DT 02-AUG-2002 (first entry)  
XX  
DE Human colon cancer related nucleotide sequence SEQ ID NO:3687.  
XX  
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200229086-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 02-OCT-2001; 2001WO-US30732.  
XX  
PR 02-OCT-2000; 2000US-237271P.  
XX  
PA (FARB ) BAYER CORP.  
XX  
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thiallingam A, Lewis ME;  
XX  
DR WPI; 2002-426115/45.  
XX  
PT New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell  
PT or tissue type, and in antisense therapy -  
XX  
PS Claim 1; Fig 1; 796pp; English.

XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample, and  
CC for detecting the presence or absence of a polynucleotide encoded by a  
CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridises to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence of  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists.  
XX  
SQ Sequence 603 BP; 157 A; 161 C; 162 G; 117 T; 6 other;

Query Match 33.5%; Score 100.6; DB 24; Length 603;  
Best Local Similarity 89.0%; Pred. No. 5.4e-22;  
Matches 154; Conservative 0; Mismatches 11; Indels 8; Gaps 4;  
QY 1 ACGAATCCGGACCTTGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60  
DB 394 ACGAATCCGGACCTTGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 452  
QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTT 118  
DB 453 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTT 512  
QY 119 GAT-GGAGATCAACACCA---GCGGGACTTTCTCCACACAAAGCGCTCAACCA 166  
DB 513 GATGGAGATCAACACCAAGCGGACTTTCTTCACACACAGCGCTNAACCA 565

RESULT 13  
AAS83051  
ID AAS83051 standard; cdna; 1257 BP.  
XX  
AC AAS83051;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #18855.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
DR P-PSDB; ABG18864.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX

PS	Claim 1; SEQ ID NO 18855; 103pp; English.	
XX	The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.	
CC	Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX	Sequence 1257 BP; 317 A; 311 C; 332 G; 296 T; 1 other;	
XX	Query Match 25.9%; Score 77.6; DB 23; Length 1257; Best Local Similarity 76.3%; Pred. No. 1.7e-14; Matches 229; Conservative 0; Mismatches 49; Indels 22; Gaps 10;	
QY	23 GGATATGTGGGACACTCGTAT--AGCCAACTCCGAGTGTCTG-CTGACAGCTTT-GTGA 78	
DB	450 GTATTTTGGGGACATTCGTATTAGGCCAAATTCGAGTGTGTTGACAGCTTTGGTGA 509	
QY	79 GACAGCAGGAGGCACATGCCAGCTGSAATCT--GACCTTCATGTCAGATCAACACAG 136	
DB	510 GACAGCAGGAGGCACATGCCAGCTGSAATCT--GACCTTCATGTCAGATCAACACAG 569	
QY	137 CGG----GACTTTCTTCACACAGCGCTCAACACATGTACAACTCCCGCAGCAAC---- 188	
DB	570 CAGCGGTGACTTTCTTCACACAGCGCTCAACACATGTACAACTCCCGCAGCAAC 629	
QY	189 ---CTCAGGCTCTGGAGTACTCACT--CTCAGGACTT-CTAGAGAAAGGCGTGGTGC 242	
DB	630 CTTCCAGGCCCTCTGGAGTACTCACT--CTCAGGACTT-CTAGAGAAAGGCGTGGTGC 689	
QY	243 AGCGCG-CTTGTCTGGGGGATGTGAGG-CTCAGGAGCTGATGAGTACTCGTGGTCTGG 300	
DB	690 AGCGCGCTTGTCTGGGGGATGTGAGG-CTCAGGAGCTGATGAGTACTCGTGGTCTGG 749	
XX	RESULT 14	
XX	ABV88973/c	
ID	ABV88973 standard; cDNA; 362 BP.	
XX	ABV88973;	
DT	13-DEC-2002 (first entry)	
XX	Human colon cancer related cDNA SEQ ID NO 2284.	
XX	Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene; ss.	
XX	Homo sapiens.	
XX	WO200258534-A2.	
PN	01-AUG-2002.	
XX	19-NOV-2001; 2001WO-US43704.	
XX	20-NOV-2000; 2000US-252222P.	
PR	06-FEB-2001; 2001US-267011P.	
PR	28-MAR-2001; 2001US-279670P.	
XX	10-JUL-2001; 2001US-304037P.	
XX	(CORI-) CORIXA CORP.	
PA	Stolk JA, Xu J, Chenault RA, Meagher MJ, Secretist H, King GE;	
XX	WPI; 2002-608400/65.	
XX	New isolated tumor colon polynucleotide and polypeptide, useful for the diagnosis, prevention and/or treatment of cancer, in particular colon cancer	
PS	Claim 1; SEQ ID NO 2284; 266pp + Sequence Listing; English.	
XX	The invention relates to a human colon tumour expressed polynucleotide (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii) complements of (i); (iii) at least 20 contiguous residues of (i); (iv) sequences that hybridize to (i), under moderately stringent conditions; (v) sequences having at least 75% or 90% identity to (i); or (vi) degenerate variants of (i). The compositions and methods of the present invention are useful for the diagnosis, prevention and/or treatment of cancer, particularly colon cancer. (I) can be used in gene therapy and (I) and (II) are useful in pharmaceutical compositions such as vaccines. CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX	Sequence 362 BP; 100 A; 98 C; 73 G; 91 T; 0 other;	
XX	Query Match 22.0%; Score 66; DB 24; Length 362; Best Local Similarity 100.0%; Pred. No. 5.2e-11; Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	235 CTTGTGTGAGCGCGCTTCTGGGGGATGTGAGCGCTCAGGAGCTGATGAGTACTCGTGG 29	
DB	362 CTTGTGTGAGCGCGCTTCTGGGGGATGTGAGCGCTCAGGAGCTGATGAGTACTCGTGG 30	
QY	295 TTCTGG 300	
DB	302 TTCTGG 297	
XX	RESULT 15	
XX	ABK45131/c	
ID	ABK45131 standard; cDNA; 362 BP.	
XX	ABK45131;	
DT	05-JUN-2002 (first entry)	
XX	cDNA encoding colon tumour protein, SEQ ID NO 682.	
XX	Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy; gene; ss.	
XX	Homo sapiens.	
XX	WO200212328-A2.	
PN	14-FEB-2002.	
XX	31-JUL-2001; 2001WO-US24218.	
XX	03-AUG-2000; 2000US-223283P.	
PR	28-MAR-2001; 2001US-279763P.	
XX	29-JUN-2001; 2001US-302051P.	
XX	(CORI-) CORIXA CORP.	
XX	King GE, Meagher MJ, Xu J, Secretist H,	

```

XX WPI: 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
PS Claim 1: SEQ ID No 682; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK44450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
SQ Sequence 362 BP; 99 A; 98 C; 74 G; 91 T; 0 other;

Query Match      22.0%; Score 66; DB 24; Length 362;
Best Local Similarity 100.0%; Pred. No. 5.2e-11;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 CCTGGTCAGCGGCTTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGG 294
Db 362 CCTGGTCAGCGGCTTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGG 303
QY 295 TTCTGG 300
Db 302 TTCTGG 297

Search completed: October 14, 2003, 14:17:01
Job time : 159.333 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 12:04:45 ; Search time 157.333 Seconds  
(without alignments)  
5147.234 Million cell updates/sec

Title: US-09-854-124-6  
Perfect score: 300  
Sequence: 1 aattccgtgtctgtcgtga.....cgaaagaagaacttttac 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_15Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	300	100.0	300	24: ABQ78751	Differentially exp
2	284.8	94.9	603	24: ABQ59992	Human colon cancer
3	284.8	94.9	858	22: AAH14111	Human CDNA sequenc
4	284.8	94.9	1171	25: ACA54679	Human NF-kappaB as
5	284.8	94.9	1210	21: AAC77820	Human cancer assoc
6	283	94.3	667	22: AAH05624	Human CDNA clone (
7	281.2	93.7	649	24: ABQ59972	Human colon cancer
8	273.8	91.3	1190	23: AAS87106	DNA encoding novel

9	231.4	77.1	1257	23	AAS83051	DNA encoding novel
10	223.4	74.5	454	24	ABL68902	Kidney cancer rela
11	102.8	34.3	678	23	AAS87103	DNA encoding novel
12	102.8	34.3	678	23	AAS87105	DNA encoding novel
13	90	30.0	1499	23	AAS87104	DNA encoding novel
14	85.2	28.4	612	23	ABL19607	Drosophila melanog
15	85.2	28.4	2670	23	ABL19606	Drosophila melanog
16	85.2	28.4	4904	23	ABL03636	Drosophila melanog
17	46.8	15.6	525	25	ABZ53276	Aspergillus oryzae
18	37.2	12.4	29559	23	AAS59546	Propionibacterium
19	36	12.0	2697	24	AAS17590	DNA encoding novel
20	34.8	11.6	1018	16	AAO2326	Human cardiac hype
21	34.8	11.6	1018	18	AAO7014	Human cardiocroph
22	34.8	11.6	1539	20	AAH87267	cDNA clone encodin
23	34.8	11.6	1539	21	AAA99914	cDNA encoding huma
24	34.8	11.6	1539	21	AAA72705	Human cardiocroph
25	34.8	11.6	1539	21	AAA72706	Human cardiocroph
26	34.8	11.6	1539	21	AAA46936	cDNA encoding nove
27	34.8	11.6	1539	22	AAO90578	Human PRO882 cDNA
28	34.8	11.6	1539	24	ABL63697	Breast cancer rela
29	34.8	11.6	3161	24	AAO49189	Porcine CD 151 gen
30	34.8	11.6	61313	23	AAS59545	Propionibacterium
31	34.6	11.5	1626	25	ABT31957	High-affinity phos
32	34.6	11.5	1626	25	ABT32039	High-affinity OSPH
33	34.2	11.4	277	24	ABN23362	Human ORFX polynuc
34	33.8	11.3	3899	24	ABO199205	Mouse ischaemic co
35	33.6	11.2	936	24	AAO17241	Streptomyces clavu
36	33.6	11.2	3773	23	ABL11845	Drosophila melanog
37	33.6	11.2	17951	23	ABL11844	Drosophila melanog
38	33.4	11.1	1118	24	ABK75437	Bacillus lichenifo
39	33	11.0	2000	24	ABO17573	Arabidopsis thalia
40	33	11.0	2577	24	AAO41592	HIV-1 subtype C is
41	33	11.0	3981	24	AAO167898	Nucleotide sequenc
42	33	11.0	4288	22	AAO20868	HIV-1 subtype (C/B
43	33	11.0	4341	24	AAO167897	Nucleotide sequenc
44	33	11.0	4343	24	AAO167896	Nucleotide sequenc
45	32.6	10.9	480	20	AAO01141	M. tuberculosis an

ALIGNMENTS

RESULT 1

ABQ78751  
ID ABQ78751 standard; cDNA; 300 BP.

XX AC ABQ78751;

XX DT 05-DEC-2002 (first entry)

XX DE Differentially expressed gene RTA00000618F.p.24.1.

XX KW Differentially expressed gene; cancer; breast cancer; colon cancer;  
KW lung cancer; prostate cancer; expressed sequence tag; EST;  
KW heat shock factor binding protein; CGI-122 gene; quiescin Q6;  
KW moderately-differentiated endometrial adenocarcinoma; TGIF protein;  
KW treacher collins syndrome protein; annexin IV; cyclophilin C;  
KW MHC class I lymphocyte antigen; HLA-E class I mRNA;  
KW glomerulosclerosis gene; 265 proteasome subunit p55;  
KW gamma interferon inducible protein; cyclin-dependent protein kinase;  
KW mitochondrial dodecenoyl-CoA delta-isomerase;  
KW serine hydroxymethyltransferase;  
KW DNA-damage-inducible RNA binding protein; ss.

XX OS Homo sapiens.

XX PN US2002076735-A1.

XX PD 20-JUN-2002.

XX PF 10-MAY-2001; 2001US-0854124.

XX PR 25-SEP-1998; 98US-101900P.

```

22-SEP-1999;      99US-0400947.

(WILL// WILLIAMS L T.
PA PA ESCOBO J.
PA PA (INNI// INNIS M A.
PA PA (GARC// GARCIA P D.
PA PA (SUDDH// SUDDUTH-KLINGER J.
PA PA (REIN// REINHARD C.
PA PA (GIES// GIESE K.
PA PA (RAND// RANDAZZO F.
PA PA (KENN// KENNEDY G C.
PA PA (POTD// POT D.
PA PA (KASS// KASSAM A.
PA PA (LAMS// LAMSON G.
PA PA (DRMA// DRMANAC R.
PA PA (CRKV// CRKVENJAKOV R.
PA PA (DICK// DICKSON M.
PA PA (DRMA// DRMANAC S.
PA PA (LABA// LABAT I.
PA PA (LESH// LESHKOWITZ D.
PA PA (KITA// KITA D.
PA PA (GARC// GARCIA V.
PA PA (JONE// JONES L W.
PA PA (STAC// STACHE-CRAIN B.
XX XX
XX XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
XX XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX XX
WPI; 2002-673346/72.

Detecting cancerous breast, colon, lung and prostate cells and
assessing metastatic potential of cells of the breast, colon and lung,
and preventing metastasis in cancer cells using nucleic acids -
XX XX
Claim 1; Page 28-29; 38pp; English.

ABQ78746-82 represents genes that are differentially expressed in
cancerous cells. Detection of the expression level of these genes, and
then comparing with the level of expression in a control cell, can be
used for detecting a cancerous breast, colon, lung or prostate cell, and
for assessing the metastatic potential of that cell. ABQ78746-47,
CC CC ABQ78753-54 and ABQ78756 correspond to an expressed sequence tag (EST);
CC CC ABQ78748-49 correspond to a heat shock factor binding protein;
CC CC ABQ78750-52 correspond to CGI-122 gene; ABQ78755 corresponds to a cDNA
sequence isolated from moderately-differentiated endometrial
adenocarcinoma; ABQ78757 corresponds to quiescin Q6; ABQ78758
corresponds to teacher collins syndrome protein; ABQ78759 corresponds
to annexin IV; ABQ78760 corresponds to TGIF protein; ABQ78761
corresponds to MHC class I lymphocyte antigen; ABQ78762 corresponds to
HLA-E class I mRNA; ABQ78763 corresponds to glomerulosclerosis gene;
CC CC ABQ78764-66 correspond to cyclophilin C; ABQ78767 corresponds to 26S
proteasome subunit p35; ABQ78768-72 corresponds to gamma interferon
inducible protein; ABQ78773 corresponds to Na+/H-exchange regulatory
co-factor; ABQ78774-77 corresponds to mitochondrial dodecenoyl-CoA
delta-isomerase; ABQ78778 corresponds to cyclin-dependent protein
kinase; ABQ78779-81 correspond to serine hydroxymethyltransferase; and
CC CC ABQ78782 corresponds to DNA-damage-inducible RNA binding protein. The
method is used to detect monitor and treat cancer of the breast, colon,
lung and prostate.
XX XX
Sequence 300 BP; 72 A; 75 C; 87 G; 66 T; 0 other;

Query Watch          100.0%; Score 300; DB 24; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.4e-82;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1    AATTCCGGTTGCTGCGGTAGAGCCTCGSCCTGCAGCTCGCGCGCATGACACTGCCGA 60
Db       |||||||
Db       1    AATTCCGGTTGCTGCGGTAGAGCCTCGGCTTCGAGCTCGCGCGCATGACACTGCCGA 60
QY      61   GTCGCAATTCCTCGCCGAGAAGAGCTGGTGTACCATTTATCCCCAACTTCAGTCTGCACAA 120

```

Matches	286;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	13	GTGGTGAGGCTCTGGGCTCGAGCTCGCGCGCCATGAGACGCTGCCGAGGTGCGAATTCCT	72						
Db	29	GACCGTGAGGCTCTGGGCTCGAGCTCGCGCGCCATGAGACGCTGCCGAGGTGCGAATTCCT	88						
QY	73	CGCCGAGAAGAGCTGGTTACCATATGCCCAACTTCAGTCTGACAAAGATCTACCCAT	132						
Db	89	CGCCGAGAAGAGCTGGTTACCATATGCCCAACTTCAGTCTGACAAAGATCTACCCAT	148						
QY	133	CGGGGGGAGCTGGGGCTTTTAAACCTGTGTTACCCGTGGAAGTGCCTGTGGCTGGC	192						
Db	149	CGGGGGGAGCTGGGGCTTTTAAACCTGTGTTACCCGTGGAAGTGCCTGTGGCTGGC	208						
QY	193	GATTAACTTAAACAAGACAGAAATGCGCTGCTCCCTCCAGAGTGGATGGATGTAGA	252						
Db	209	GATTAACTTAAACAAGACAGAAATGCGCTGCTCCCTCCAGAGTGGATGGATGTAGA	268						
QY	253	AAAGTTGGAGAAGATGAGGATCATGAACAGAAAGAAACACTTTTAC	300						
Db	269	AAAGTTGGAGAAGATGAGGATCATGAACAGAAAGAAACACTTTTAC	316						

RESULT 3  
AAH14111  
ID AAH14111 standard; cDNA; 858 bp.  
XX  
XX  
AC AAH14111;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human cDNA sequence SEQ ID NO:11295.  
XX  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

RESULT 4	
ACA54679	
ID	ACA54679 standard; cDNA; 1171 BP.
XX	
XX	
ACA54679;	
XX	
XX	
05-JUN-2003	(first entry)
DT	
XX	
DE	Human NF-kappaB associated polynucleotide sequence #119.

RESULT 4  
ACA54679  
ID ACA54679 standard; cDNA: 1171 BP.

XX 31-OCT-2002.  
 XX 19-APR-2002; 2002WO-US12636.  
 XX 19-APR-2001; 2001US-284962P.  
 XX 26-APR-2001; 2001US-286645P.  
 XX 09-JAN-2002; 2002US-346986P.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Carman J, Feder J, Nadler S;  
 XX WPI; 2003-093119/08.  
 XX P-PSDB; AB069618.  
 XX Novel NF-kappaB-associated polypeptides and polynucleotides useful for  
 XX diagnosing, treating and preventing cancer, hepatic disorders, aberrant  
 XX apoptosis, viral infections, autoimmune disorders, asthma and stroke -  
 XX Claim 4; Fig 16; 608pp; English.  
 XX The present invention relates to the isolation of human nuclear  
 XX factor-kappaB (NF-kappaB) associated polypeptides and polynucleotides.  
 XX The NF-kappaB associated polypeptide and polynucleotide sequences  
 XX are useful for preventing, treating or ameliorating various disorders  
 XX including immune disorders, inflammatory disorders, cancers,  
 XX disorders relating to aberrant apoptosis, hepatic disorders,  
 XX Hodgkin's lymphomas, haematopoietic tumours, hyper-IGM syndromes,  
 XX hypohidrotic ectodermal dysplasia, X-linked anhidrotic ectodermal  
 XX dysplasia, immunodeficiency, al incontinentia pigmenti, viral  
 XX infections (e.g. those caused by human immunodeficiency virus (HIV),  
 XX human T-cell lymphotropic virus (HTLV), hepatitis B, hepatitis C,  
 XX Epstein Barr virus (EBV), influenza), rheumatoid arthritis,  
 XX inflammatory bowel disease, colitis, asthma, atherosclerosis, cachexia,  
 XX euthyroid sick syndrome, stroke, experimental allergic encephalomyelitis,  
 XX disorders related to aberrant acute phase responses, hypercongenital  
 XX conditions, birth defects, necrotic lesions, wounds, organ transplant  
 XX rejection, disorders related to aberrant signal transduction,  
 XX hyperproliferative disorders, diseases of the pancreas (e.g. diabetes  
 XX mellitus, vitamin B12 malabsorption), neurological disorders (e.g.  
 XX Huntington's chorea), Turner's syndrome, bacterial infections, e.g.  
 XX cardiovascular disorders, infertility, psoriasis and haemolytic anaemia.  
 XX The present sequence represents a human NF-kappaB associated  
 XX polynucleotide of the invention.  
 XX Sequence 1171 BP; 296 A; 298 C; 309 G; 268 T; 0 other;  
 XX  
 XX Query Match 94.9%; Score 284.8; DB 25; Length 1171;  
 XX Best Local Similarity 99.3%; Pred. No. 1.2e-77;  
 XX Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 13 GTCGGTGAGGCTGTGGCTGCGCTGCGCGCCATGGAGCTGCGGAGTCAATTCCT 72  
 Db 43 GACCGTGAGGCTGTGGCTGCGCTGCGCGCCATGGAGCTGCGGAGTCAATTCCT 102  
 QY 73 CGCCGAGAAGGAGCTGGTTACCAATATCCCACTTCACTGCGAAGATCACTCAT 132  
 Db 103 CGCCGAGAAGGAGCTGGTTACCAATATCCCACTTCACTGCGAAGATCACTCAT 162  
 QY 133 CGGGGGGACCTGGGGCCCTTTAACTGTTTACCCGTGGAGTGGCCCTGTGGCTGGC 192  
 Db 163 CGGGGGGACCTGGGGCCCTTTAACTGTTTACCCGTGGAGTGGCCCTGTGGCTGGC 222  
 QY 193 GATTACCTGAACAAAGACAAATGTGCGCTGCTCCCTCCAGATGGATGGATGAGA 252  
 Db 223 GATTACCTGAACAAAGACAAATGTGCGCTGCTCCCTCCAGATGGATGGATGAGA 282  
 QY 253 AAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAGAACTTTTAC 300  
 Db 283 AAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAGAACTTTTAC 330

RESULT 5  
 AAC77820  
 ID AAC77820 standard; cDNA; 1210 BP.  
 XX  
 AC AAC77820;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human cancer associated gene sequence SEQ ID NO:214.  
 XX  
 KW Human; cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;  
 KW antidiabetic; antisthmatic; antirheumatic; antiallergic; antiviral;  
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055350-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05882.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2000-587533/55.  
 DR P-PSDB; AAB43611.  
 XX  
 PT Novel isolated nucleic acids comprising sequences encoding peptides  
 XX useful for treating or diagnosing e.g. cancer -  
 PS Claim 1; Page 785-786; 2352pp; English.  
 XX  
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnary; immunomodulator;  
 CC antidiabetic; antisthmatic; antirheumatic; antiallergic; antiviral;  
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; coagulant;  
 CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 1210 BP; 320 A; 300 C; 320 G; 268 T; 2 other;  
 XX  
 XX Query Match 94.9%; Score 284.8; DB 21; Length 1210;  
 XX Best Local Similarity 99.3%; Pred. No. 1.2e-77;  
 XX Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 13 GTCGGTGAGGCTGTGGCTGCGCTGCGCGCCATGGAGCTGCGGAGTCAATTCCT 72



Db 47 GACGTGAGGCTCTGGCCCTGACGCTCGCCGCCCATGAGCGCTGCGAGGTGCAATTCCT 106  
QY 73 CGCCGAGAAGGAGCTGTTTACCATTATCCCAACTTACGTCGACAAAGATCTACCTCAT 132  
Db 107 CGCCGAGAAGGAGCTGTTTACCATTATCCCAACTTACGTCGACAAAGATCTACCTCAT 166  
QY 133 CGGGGGGAGCCTGGGGCCCTTTAACCTGGTTTACCGTGGGAAGTCCCTGTGGCTGGC 192  
Db 167 CGGGGGGAGCCTGGGGCCCTTTAACCTGGTTTACCGTGGGAAGTCCCTGTGGCTGGC 226  
QY 193 GATTAACCTGAAACAAAGACAGAAATGTCGCTGCTCCCTCCAGAGTGGATGATGTACA 252  
Db 227 GATTAACCTGAAACAAAGACAGAAATGTCGCTGCTCCCTCCAGAGTGGATGATGTACA 286  
QY 253 AAAGTTGGAGAAGATGAGGAGTCATGACGAAGGAAGAACTTTTAC 300  
Db 287 AAAGTTGGAGAAGATGAGGAGTCATGACGAAGGAAGAACTTTTAC 334  
RESULT 6  
AAH05624  
ID AAH05624 standard; cDNA; 667 BP.  
XX  
AC AAH05624;  
XX  
XX 26-JUN-2001 (first entry)  
DT  
DE Human cDNA clone (5'-primer) SEQ ID NO:2459.  
DE  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
PN  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
full-length cDNAs defined in the specification, and for the detection  
and/or diagnosis of the abnormality of the proteins encoded by the  
full-length cDNAs -  
XX  
XX Claim 1; SEQ ID 2459; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
full-length cDNAs defined in the specification. Where a primer set  
comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination of  
the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 667 BP; 171 A; 176 C; 177 G; 137 T; 6 other;  
Query Match 94.3%; Score 283; DB 22; Length 667;  
Best Local Similarity 100.0%; Pred. No. 3.5e-77;  
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 TGAGGCTCTGGCCTGCGAGCTCGCGCGCCATGACGCTGCGAGGTGCAATTCCTCGCGG 77  
Db 7 TGAGGCTCTGGCCTGCGAGCTCGCGCGCCATGACGCTGCGAGGTGCAATTCCTCGCGG 66  
QY 78 AGAAGGAGCTGTTTACCATTATCCCAACTTACGTCGACAAAGATCTACCTCATCGGG 137  
Db 67 AGAAGGAGCTGTTTACCATTATCCCAACTTACGTCGACAAAGATCTACCTCATCGGG 126  
QY 138 GGGACCTGGGGCCCTTTTAAACCTGGTTTACCCTGGTGGAAAGTGGCCCTGTGGCGGATTA 197  
Db 127 GGGACCTGGGGCCCTTTTAAACCTGGTTTACCCTGGTGGAAAGTGGCCCTGTGGCGGATTA 186  
QY 198 ACCTGAAACAAAGACAGAAATGTCGCTGCTCCCTCCAGAGTGGATGGATAGAAAAGT 257  
Db 187 ACCTGAAACAAAGACAGAAATGTCGCTGCTCCCTCCAGAGTGGATGGATAGAAAAGT 246  
QY 258 TGGAGAAGATGAGGGATCATGACGAAGGAAGAACTTTTAC 300  
Db 247 TGGAGAAGATGAGGGATCATGACGAAGGAAGAACTTTTAC 289  
RESULT 7  
ABQ59972  
ID ABQ59972 standard; cDNA; 649 BP.  
XX  
AC ABQ59972;  
XX  
DT 02-AUG-2002 (first entry)  
DE Human colon cancer related nucleotide sequence SEQ ID NO:3667.  
DE  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200229086-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 02-OCT-2001; 2001WO-US30732.  
XX  
PR 02-OCT-2000; 2000US-237271P.  
XX  
PA (FARB ) BAYER CORP.  
XX  
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thialingam A, Lewis ME;  
XX  
WPI; 2002-426115/45.  
XX  
XX New isolated nucleic-acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell  
PT or tissue type, and in antisense therapy -  
XX  
PS Claim 1; Fig 1; 796pp; English.  
XX

ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins encoded by the ABQ60776 to ABQ60778 nucleic acid sequences. (I) can be used in antisense therapy. An antibody immunoreactive with a polypeptide encoded by (I) is useful for detecting cancer in a patient sample, and for detecting the presence or absence of a polynucleotide encoded by a nucleic acid which hybridises to (I) in a cell. A probe/primer derived from (I) can be used for determining the presence of a nucleic acid which hybridises to (I), and for determining the phenotype of cells in a sample of cells from a patient. (I) is useful for determining the presence of colon cancer in a cell or tissue type, for determining the presence or state of other type of cancer, in antisense therapy, to generate macroarrays on a solid surface, to identify a chromosome on which the corresponding gene resides, and in tissue profiling, forensics, genetic analysis, mapping and diagnostic applications. (I) can be used to raise antibodies, and to screen for peptide analogues and antagonists.

Sequence 649 BP; 159 A; 163 C; 169 G; 138 T; 20 other:  
SQ

Query Match	93.7%;	Score 281.2;	DB 24;	Length 649;
Best Local Similarity	98.3%;	Pred. No. 1.2e-76;		
Matches 283;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
QY	13	GTCCGTGAGGCTCTGGCCCTGCAGCTCGCGCGGCATGGACGCTGCCGAGGTGCGAATTCCT	72	
Db				
	32	GACCTGTAGGCTTTGGCCCTGCATCTCGCGCGGCATGGACGCTGCCGAGGTGCGAATTCCT	91	
QY	73	CGCCGAGGAGGCTGGTTACCATATATCCCACTTCAGTCTGGACAAGATCTACCTCAT	132	
Db				
	92	CGCCGAGGAGGCTGGTTACCATATATCCCACTTCAGTCTGGACAAGATCTACCTCAT	151	
QY	133	CGGGGGGACCTCGGGGCCCTTTTAACTCTGGTTTACCGTGGAAAGTCCCCCTGTGGCTGGC	192	
Db				
	152	CGGGGGGACCTCGGGGCCCTTTTAACTCTGGTTTACCGTGGAAAGTCCCCCTGTGGCTGGC	211	
QY	193	GATTAACCTGAAACAAGACAGAAATCTCGCCTGCTCCCTCCACAGATGGATGGATCTAGA	252	
Db				
	212	GATTAACCTGAAACAAGACAGAAATCTCGCCTGCTCCCTCCACAGATGGATGGATCTAGA	271	
QY	253	AAAGTTTGAGAGATGAGGAGTCATGACGAAAGGAGAACTTTTAC	300	
Db				
	272	AAAGTTTGAGAGATGAGGAGTCATGACGAAAGGAGAACTTTTAC	319	

RESULT 8  
AAS87106  
ID AAS87106 standard; cDNA; 1190 BP.  
XX  
XX  
AAS87106;  
XX  
XX  
13-FEB-2002 (first entry)  
XX  
XX  
DNA encoding novel human diagnostic protein #22910.  
XX  
XX  
Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

AA	WO200175067-A2.
PN	11-OCT-2001.
XX	30-MAR-2001; 2001WO-US08631.
XX	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	(HYSE-) HYSEQ INC.
XX	Drmanac RT, Liu C, Tang YT;
PI	WPI; 2001-639362/73.
DR	

DR P-PSDB; ABG22919.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 1; SEQ ID No 22910; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

SQ Sequence 1190 BP; 298 A; 304 C; 310 G; 278 T; 0 other;

Query Match	91.3%	Score 273.8	DB 23	Length 1190
Best Local Similarity	99.0%	Pred. No. 3e-74		
Matches 286	Conservative 0	Mismatches 2	Indels 1	Gaps 1
QY	13	GTCCGCTGAGGCTCTGGCCTCTCAGCTCGCGCGCCGCAATGGAGCGTCGCCAGGTGCGAATTCCT	72	
Db	43	GACCGTGAGGCTCTGGCCTCTCAGCTCGCGCGCCGCAATGGAGCGTCGCCAGGTGCGAATTCCT	102	
QY	73	CGCCGAGAAGGACCTGGTTACCATATATCCCAACTTTCAGTCTGGACAAGATCTACCTCAT	132	
Db	103	CGCCGAGAAGGACCTGGTTACCATATATCCCAACTTTCAGTCTGGACAAGATCTACCTCAT	162	
QY	133	CGGGGGGGACCTGGGGCCCTTTTAAACCTTGTTTTACCGCTGGAAGTGCCTCTGGGTGGC	192	
Db	163	CGGGGGGGACCTGGGGCCCTTTTAAACCTTGTTTTACCGCTGGAAGTGCCTCTGGGTGGC	222	
QY	193	GATTAACTTGAAACAAGACAGAAATCTGCCCTGCTCCCTCCAGATGGATGATGTAGA	252	
Db	223	GATTAACTTGAAACAAGACAGAAATCTGCCCTGCTCCCTCCAGATGGATGATGTAGA	282	
QY	253	AAAGTTGGAGAGATGAGGATCATGAA - CGAAAGGAAGAAACCTTTTAC	300	
Db	283	AAAGTTGGAGAGATGAGGATCATGAA - CGAAAGGAAGAAACCTTTTAC	331	

RESULT 9	
AAS83051	
ID	AAS83051 standard; cDNA; 1257 BP.
XX	
XX	
XX	AAS83051;
XX	
XX	
XX	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #18855.
DE	
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.



The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour.

Claim 1; SEQ ID No 7239; 44pp; English.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 1; SEQ ID No 22907; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 454 BP; 128 A; 113 C; 116 G; 97 T; 0 other;

Query Match 74.5%; Score 223.4; DB 24; Length 454;  
Best Local Similarity 99.6%; Pred. No. 7.4e-59;  
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 76 CGAGAGGACGTGGTTACCATTATCCCAACTTCAGCTGGACAGAGTCTACCTCATCGG 135  
DB 2 CGATAAGGACGTGGTTACCATTATCCCAACTTCAGCTGGACAGAGTCTACCTCATCGG 61

QY 136 GGGGGACCTGGGCCCTTTTAACCCCTGGTTTACCCGTGGAAAGTGCCCCCTGGCTGGCGAT 195  
DB 62 GGGGGACCTGGGCCCTTTTAACCCCTGGTTTACCCGTGGAAAGTGCCCCCTGGCTGGCGAT 121

QY 196 TAACCTGAACAAGAAGACAGAAATGCGCTGCCTCCCTCAGAGTGGATGTAGAAAAA 255  
DB 122 TAACCTGAACAAGAAGACAGAAATGCGCTGCCTCCCTCAGAGTGGATGTAGAAAAA 181

QY 256 GTTGGAAGAGTAGGGATCATGAACGAAGGAAGAACTTTTAC 300  
DB 182 GTTGGAAGAGTAGGGATCATGAACGAAGGAAGAACTTTTAC 226

RESULT 11  
AAS87103  
ID AAS87103 standard; cDNA; 678 BP.  
XX AC  
XX AAS87103;  
XX DT  
XX 13-FEB-2002 (first entry)  
XX DE  
XX DNA encoding novel human diagnostic protein #22907.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX OS Homo sapiens.  
OS WO200175067-A2.  
PN PN  
PD PD  
PF 11-OCT-2001.  
PP 30-MAR-2001; 2001WO-US08631.  
PX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
PX XX  
PA (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;

chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set -

Claim 1; SEQ ID 7239; 44pp; English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour.

Claim 1; SEQ ID No 22907; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 678 BP; 177 A; 185 C; 206 G; 110 T; 0 other;

Query Match 34.3%; Score 102.8; DB 23; Length 678;  
Best Local Similarity 86.9%; Pred. No. 1.5e-21;  
Matches 113; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 13 GTGCGTGAAGCTCTGGCTGCAGCTCGCGCCGCGCATGGAGCTGCCGAGTGGCGAGTTCCT 72  
DB 255 GACTGTGAGGCCGAGCTGGCCACTCGCGCCGCGCATGGAGTGGCGAGTTCCT 314

QY 73 CGCGGAGAAGAGCTGGTTACCATTATCCCCTAACCTTCAGTCTGGACAAGATCTACCTCAT 132  
DB 315 GGCTGAGAAGAGCTGGTTACCATTATCCCCTAACCTTCAGTCTGGACAAGATCTACCTCAT 374

QY 133 CGGGGGGGAC 142  
DB 375 TGGGGTGGAC 384

RESULT 12  
AAS87105  
ID AAS87105 standard; cDNA; 678 BP.  
XX AC  
XX AAS87105;  
XX DT  
XX 13-FEB-2002 (first entry)  
XX DE  
XX DNA encoding novel human diagnostic protein #22909.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX OS Homo sapiens.  
OS WO200175067-A2.  
PN PN  
PD PD  
PF 11-OCT-2001.  
PP 30-MAR-2001; 2001WO-US08631.  
PX 30-MAR-2001; 2001WO-US08631.  
PI



OS Drosophila melanogaster.  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1: SEQ ID NO 10294; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 612 BP; 155 A; 166 C; 138 G; 138 T; 0 other;  
 Query Match 28.4%; Score 85.2; DB 23; Length 612;  
 Best Local Similarity 60.3%; Pred. No. 3.8e-16;  
 Matches 141; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
 QY 47 ATGGACGCTCGCGAGGTCGAATTCCTCGCGAGAGGAGGCTGTACCATTTATCCCCAAC 106  
 DB 1 ATGGATCCTTCAATTTATGATTTATTTGGCGAAAATGCATGATCAGCATAAATACCGAAC 60  
 QY 107 TTCAGTCTGGACAAGATCTACCTCATCGGGGGGACCTGGGGCCCTTTTAAACCTTGGTTTA 166  
 DB 61 TTCAGCAACGAGCCTCTGCACCTGATATACGACCTGTGGGCCCTTTCCGAGCGGTTT 120  
 QY 167 CCCGTGGAAGTCCCTGTGGCTGGCGATTAACCTGAACAAAGACAGAAATGTCGCTG 226  
 DB 121 CCCGTCTTGGTCCCTGTGGATGGCCACGCTGTGCGCAAGCAACAAAAGTCCGGAAT 180  
 QY 227 CTCCTCCAGATGGATGGATGAGAAAAGTTGGAGAAGATGAGGGATCATGAA 280  
 DB 181 GTACCTCCAGATGGATGGATGATATATTTGGAGGAATCAAGGAGGAGAA 234

RESULT 15  
 ABL19606  
 ID ABL19606 standard; DNA; 2670 BP.  
 XX  
 AC ABL19606;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 10291.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX

PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1: SEQ ID NO 10291; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2670 BP; 733 A; 627 C; 635 G; 675 T; 0 other;  
 Query Match 28.4%; Score 85.2; DB 23; Length 2670;  
 Best Local Similarity 60.3%; Pred. No. 7e-16;  
 Matches 141; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
 QY 47 ATGGACGCTCGCGAGGTCGAATTCCTCGCGAGAGGAGGCTGTACCATTTATCCCCAAC 106  
 DB 1001 ATGGATCCTTCAATTTATGATTTATTTGGCGAAAATGCATGATCAGCATAAATACCGAAC 1060  
 QY 107 TTCAGTCTGGACAAGATCTACCTCATCGGGGGGACCTGGGGCCCTTTTAAACCTTGGTTTA 166  
 DB 1061 TTCAGCAACGAGCCTCTGCACCTGATATACGACCTGTGGGCCCTTTCCGAGCGGTTT 1120  
 QY 167 CCCGTGGAAGTCCCTGTGGCTGGCGATTAACCTGAACAAAGACAGAAATGTCGCTG 226  
 DB 1121 CCCGTCTTGGTCCCTGTGGATGGCCACGCTGTGCGCAAGCAACAAAAGTCCGGAAT 1180  
 QY 227 CTCCTCCAGATGGATGGATGAGAAAAGTTGGAGAAGATGAGGGATCATGAA 280  
 DB 1181 GTACCTCCAGATGGATGGATGATATATTTGGAGGAATCAAGGAGGAGAA 1234

Search completed: October 14, 2003, 14:17:03  
 Job time : 159.333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 13:05:16 ; Search time 36.6667 Seconds  
(without alignments)  
3611.317 Million cell updates/sec

Title: US-09-854-124-6

Perfect score: 300

Sequence: 1 aatccgtgtctgtcgtga.....cgaaaggagaacattttac 300

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38	12.7	1335	4	US-09-252-991A-5087
C 2	38	12.7	1389	4	US-09-252-991A-5133
C 3	34.8	11.6	1018	1	US-08-444-083-6
C 4	34.8	11.6	1018	1	US-08-286-304-6
C 5	34.8	11.6	1018	1	US-08-442-745-6
C 6	34.8	11.6	1018	1	US-08-443-129-6
C 7	34.8	11.6	1018	1	US-08-443-952-6
C 8	34.8	11.6	1018	1	US-08-443-130-6
C 9	34.8	11.6	1018	3	US-08-898-911-6
C 10	34.8	11.6	1018	5	PCT-US95-04467-6
C 11	34.8	11.6	1539	4	US-09-648-183-1
C 12	34.8	11.6	1539	4	US-09-648-183-2
C 13	33.6	11.2	429	4	US-09-252-991A-5055
C 14	33	11.0	711	4	US-09-252-991A-13138
C 15	33	11.0	783	4	US-09-252-991A-12804
C 16	32.6	10.9	480	4	US-09-073-009-135
C 17	32.6	10.9	1709	4	US-09-594-193-12
C 18	32.2	10.7	4403765	3	US-09-103-840A-2
C 19	32.2	10.7	4411529	3	US-09-103-840A-1
C 20	32	10.7	4403765	3	US-09-103-840A-2
C 21	32	10.7	4411529	3	US-09-103-840A-1
C 22	31.6	10.5	486	4	US-09-252-991A-6292
C 23	31.6	10.5	564	4	US-09-252-991A-6455
C 24	31.6	10.5	678	4	US-09-252-991A-2321
C 25	31.6	10.5	753	4	US-09-252-991A-2382
C 26	31.6	10.5	1241	1	US-08-471-033-39
C 27	31.6	10.5	1241	1	US-08-471-033-42

28	31.6	10.5	1241	2	US-08-471-044-39	Sequence 39, Appl
29	31.6	10.5	1241	2	US-08-471-044-42	Sequence 42, Appl
30	31.6	10.5	1241	2	US-08-463-483A-39	Sequence 39, Appl
31	31.6	10.5	1241	2	US-08-463-483A-42	Sequence 42, Appl
32	31.6	10.5	1241	2	US-08-471-046A-39	Sequence 39, Appl
33	31.6	10.5	1241	2	US-08-471-046A-42	Sequence 42, Appl
34	31.6	10.5	1241	2	US-08-470-566B-39	Sequence 39, Appl
35	31.6	10.5	1241	2	US-08-470-566B-42	Sequence 42, Appl
36	31.6	10.5	1241	2	US-08-469-334-39	Sequence 39, Appl
37	31.6	10.5	1241	2	US-08-469-334-42	Sequence 42, Appl
38	31.6	10.5	1241	3	US-09-300-529-39	Sequence 39, Appl
39	31.6	10.5	1241	3	US-09-300-529-42	Sequence 42, Appl
40	31.6	10.5	1358	1	US-08-471-033-45	Sequence 45, Appl
41	31.6	10.5	1358	2	US-08-471-044-45	Sequence 45, Appl
42	31.6	10.5	1358	2	US-08-463-483A-45	Sequence 45, Appl
43	31.6	10.5	1358	2	US-08-471-046A-45	Sequence 45, Appl
44	31.6	10.5	1358	2	US-08-470-566B-45	Sequence 45, Appl
45	31.6	10.5	1358	2	US-08-469-334-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1  
US-09-252-991A-5087/c  
; Sequence 5087, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5087  
; LENGTH: 1335  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5087

Query Match	12.7%	Score 38;	DB 4;	Length 1335;
Best Local Similarity	51.8%	Pred. No. 0.061;		
Matches	86;	Conservative	0;	Mismatches 80; Indels 0; Gaps 0;
QY	10	GCTGTCGTGAGGCTCTGGCCTCGAGCTCGCGCCGCGCATGACGCTGCCGAGGTGCGAATT	69	
Db	421	GGTGGCGATGGCCTTCAGCCTCGCTGGAGATACCTGTACGGGGCGCTACTGGGGATG	362	
QY	70	CCTCGCGGAGAAAGAGCTGGTTACCATTTCCCACTTCAGTCTGGACAGATCTACCT	129	
Db	361	CCCTGGCGAGTTTCAGCGGCTCGACCTTCGAGACCTTCTTACCGGCTGGAATTCGC	302	
QY	130	CATCGGGGGGAGCTGGGGCTTTTACCTCTGGTTTACCGGTGAA	175	
Db	301	CATCGGGCGCGCTGGCGGCTTCGACGCCGCGCCGCGGCGGCGAA	256	

RESULT 2  
US-09-252-991A-5133  
; Sequence 5133, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5133  
; LENGTH: 1389  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5133

Query Match 12.7%; Score 38; DB 4; Length 1389;  
Best Local Similarity 51.8%; Pred. No. 0.062;  
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 10 GCTGTCGTGAGGCTCTGCGCTGACGCTGCGCCCGCCATGACGCTGCGGAGGTGGAATT 69  
DB 1059 GGTGCGGATGGGTTTCAGCGTGGAGATACCTGTACGGCGCCTACTCTGGGATG 1118  
QY 70 CCTCGCCGAGGAGGAGCTGGTTACCATTTATCCCAACTTACGCTGCGACAAGATCTACCT 129  
DB 1119 CTTGGCGAGTTCGACCGGCTGCACTTCGAGACCTGCTTCTACCAAGGCGCTGGAATTCG 1178  
QY 130 CATCGGGGGGACCTGCGGCGCTTTTAAACCTGTTTACCCGTGGAA 175  
DB 1179 CATCGGCGCGGCTTGGCGGCTTCGACGCGCGGCGCCAGGCGAA 1224

RESULT 3  
US-08-444-083-6  
; Sequence 6, Application US/08444083  
; Patent No. 5571675  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Joffre  
; APPLICANT: Chien, Kenneth  
; APPLICANT: King, Kathleen  
; APPLICANT: Pennica, Diane  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/444,083  
; FILING DATE: 17-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/233609  
; FILING DATE: 25-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/286304  
; FILING DATE: 05-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: 894PID5  
; TELEPHONE: 415/225-8674  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1018 bases

; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-444-083-6  
Query Match 11.6%; Score 34.8; DB 1; Length 1018;  
Best Local Similarity 50.6%; Pred. No. 0.49;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 21 GGCTCTGCGCTGACGCTGCGCGCCGATGAGGAGCTGCGGAGGTGCAAAATCTCGCGCGAGA 80  
DB 484 GGGCGGAGCCCCCGCCGCCCTCAGCGGCTCGGCCACCGGGGTCTTCCCGCGCA 543  
QY 81 AGGAGCTGGTTACCATTTATCCCAACTTTCAGTCTGGACAAGATCTACCTCATTCGCGGGGG 140  
DB 544 AGGTGCTGGGCTCGCGCTTTGGCGGCTTACCGCGAGTGCTGAGCGCGCAGCGAGGGG 603  
QY 141 ACCTGGGCGCTTTTAAACCTGTTTACCGGTGTTTACCGTGGAAAGTGCCTCTG 186  
DB 604 ACCTGGGCGAGCTGCTGCGCGGGGCTCGGCTGAGCGCGCGGGG 649

RESULT 4  
US-08-286-304-6  
; Sequence 6, Application US/08286304  
; Patent No. 5571893  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Joffre  
; APPLICANT: Chien, Kenneth  
; APPLICANT: King, Kathleen  
; APPLICANT: Pennica, Diane  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286,304  
; FILING DATE: 05-AUG-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/233609  
; FILING DATE: 25-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hasak, Janet E.  
; REGISTRATION NUMBER: 28,616  
; REFERENCE/DOCKET NUMBER: 894P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1896  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1018 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-286-304-6

Query Match 11.6%; Score 34.8; DB 1; Length 1018;  
Best Local Similarity 50.6%; Pred. No. 0.49;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;





Db 604 ACCTGGGCGACCTGCTGCCCGGGGCTCGGCTGAGCGCGGGG 649

## RESULT 7

US-08-443-952-6

; Sequence 6, Application US/08443952

; Patent No. 5679545

; GENERAL INFORMATION:

; APPLICANT: Baker, Joffre

; APPLICANT: Chien, Kenneth

; APPLICANT: King, Kathleen

; APPLICANT: Pennica, Diane

; APPLICANT: Wood, William

; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/443,952

; FILING DATE: 17-MAY-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/233609

; FILING DATE: 25-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/286304

; FILING DATE: 05-AUG-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Torchia, Timothy E.

; REGISTRATION NUMBER: 36,700

; REFERENCE/DOCKET NUMBER: 894PID4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-8674

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1018 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-443-952-6

Query Match 11.6%; Score 34.8; DB 1; Length 1018;  
Best Local Similarity 50.6%; Pred. No. 0.49;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 21 GGCTCTGGCTGCGAGCTGCGCGCCCATGAGCGTGGCGAGTGCAGATTCTCGCCGAGA 80

Db 484 GGGCGAGCCCCCGCCGCCACCGCTCAGCCGCTCGGCCACCGGGTCTTCCCGCCA 543

Qy 81 AGGAGCTGGTTTACCATTTATCCCAACTTACGCTTGGACAAGATCTACCTCATCGGGGGG 140

Db 544 AGTGCTGGGGCTCCGGCTTTGCGGCTCTACCCGAGTGGCTGAGCGGACCGAGGGCG 603

Qy 141 ACCTGGGGCCTTTTAACTGGTTTACCGTGGAGTGCCCTGTG 186

Db 604 ACCTGGGGCCACTGCTGCCCGGGGGCTCGGCTGAGCGCGCGGGG 649

## RESULT 8

US-08-443-130-6

; Sequence 6, Application US/08443130

; Patent No. 5723585

; GENERAL INFORMATION:

; APPLICANT: Baker, Joffre

; APPLICANT: Chien, Kenneth

; APPLICANT: King, Kathleen

; APPLICANT: Pennica, Diane

; APPLICANT: Wood, William

; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/443,130

; FILING DATE: 17-MAY-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/233609

; FILING DATE: 25-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/286304

; FILING DATE: 05-AUG-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Torchia, Timothy E.

; REGISTRATION NUMBER: 36,700

; REFERENCE/DOCKET NUMBER: 894PID3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-8674

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1018 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-443-130-6

## Query Match

11.6%; Score 34.8; DB 1; Length 1018;

Best Local Similarity 50.6%; Pred. No. 0.49;

Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 21 GGCTCTGGCTGCGAGCTGCGCGCCCATGAGCGTGGCGAGTGCAGATTCTCGCCGAGA 80

Db 484 GGGCGAGCCCCCGCCGCCACCGCTCAGCCGCTCGGCCACCGGGTCTTCCCGCCA 543

Qy 81 AGGAGCTGGTTTACCATTTATCCCAACTTACGCTTGGACAAGATCTACCTCATCGGGGGG 140

Db 544 AGTGCTGGGGCTCCGGCTTTGCGGCTCTACCCGAGTGGCTGAGCGGACCGAGGGCG 603

Qy 141 ACCTGGGGCCTTTTAACTGGTTTACCGTGGAGTGCCCTGTG 186

Db 604 ACCTGGGGCCACTGCTGCCCGGGGGCTCGGCTGAGCGCGCGGGG 649

RESULT 9

US-08-898-911-6

; Sequence 6, Application US/08898911

; Patent No. 6117650

; GENERAL INFORMATION:

; APPLICANT: King, Kathleen

; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor



```

; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 1
; LENGTH: 1539
; TYPE: DNA
; ORGANISM: Human
US-09-648-183-1

```

Query Match	11.6%;	Score 34.8;	DB 4;	Length 1539;
Best Local Similarity	50.6%;	Prod. No. 0.57;		
Matches 84;	Conservative 0;	Mismatches 82;	Indels 0;	Gaps 0;
QY	21	GGCTGTGCCGTGACGTCCGGCGCGCCATGACACGCTCCGAGGTGCGAATTCCTTCGCCCGAGA	80	
Db	484	GGCGGAGCCCCCGCGCCACCGCCTCAGCCGCTCCGCCACCGGGGTCTTCCCGCGCA	543	
QY	81	AGGAGCTGGTTACCATTTATCCCAACTTCAAGTCTGGACAAGATCTACCTCATCGGGGGGG	140	
Db	544	AGGTGCTGGGCTCCGGCTTGGGGCTCTACCGCAGTGGCTGAGCCGCCACCGAGGGCG	603	
QY	141	ACCTGGGGCCCTTTAAACCCCTGGTTTTACCGTGGAACTGCCCTGTG	186	
Db	604	ACCTGGGCACTGCTGCCGGGGGCTCGGCTTGAAGCGCGGGG	649	

RESULT 12  
US-09-648-183-2/c

```

: Sequence 2, Application US/09648183
: Patent No. 6472585
: GENERAL INFORMATION:
: APPLICANT: Botstein, David
: APPLICANT: Goddard, Audrey
: APPLICANT: Lawrence, David A.
: APPLICANT: Pennica, Diane
: APPLICANT: Roy, Margaret A.
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: Cardiotrophin-1 Compositions and Methods for the
: TREATMENT OF TUMOR
: FILE REFERENCE: P2533C1
: CURRENT APPLICATION NUMBER: US/09/648,183
: CURRENT FILING DATE: 2000-08-25
: PRIOR APPLICATION NUMBER: US 09/234,730
: PRIOR FILING DATE: 1999-01-21
: PRIOR APPLICATION NUMBER: US 09/033,114
: PRIOR FILING DATE: 1998-03-02
: PRIOR APPLICATION NUMBER: US 08/733,850
: PRIOR FILING DATE: 1996-10-18
: PRIOR APPLICATION NUMBER: US 08/443,129
: PRIOR FILING DATE: 1995-05-17
: PRIOR APPLICATION NUMBER: US 08/286,304
: PRIOR FILING DATE: 1994-08-05
: PRIOR APPLICATION NUMBER: US 08/233,609
: PRIOR FILING DATE: 1994-04-25
: PRIOR APPLICATION NUMBER: US 60/113,296
: PRIOR FILING DATE: 1998-12-22
: NUMBER OF SEQ ID NOS: 6

```

	Query Match Best Local Matches	Similarity 84;	Conservative 0;	Mismatches 82;	Indels 0;	Gaps 0;
QY	21	GGCTCTGGCTTCAGCTTCGGCGCGCCATGAGACGCTCGCGAGGTCGAATTCCTCGCCGAGA	80			
dDb	1056	GGGCGGAGCGCCCCCGCGCGCACGCGCTCAGCCGCTCCGCGCGGGTCTTCCCGCGCA	997			
QY	81	AGGAGCTGGTTACCATTTATCCCACTTCAGTCTGGACAAAGATCTACCTCATCGGGGGGGG	140			
dDb	996	AGGTGCTGGGGCTCCGCGTTTGGGGCTCTACCGGAGTGGGTGAGCCGACACCGAGGGCG	937			

QY	141	ACCTGGGGCCCTTTTAAACCCCTGGTTTACCCCTGGAAGTGCCCTGTG	186
Dδ	936	ACCTGGGCCAGCTGCTGCCCCGGGGCTCGGCCTGAGCGCCGCGGGG	891

```

RESULT 13
US-09-252-991A-5055/c
; Sequence 5055, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5055
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5055

```

	Query Match	11.2%	Score 33.6;	DB 4;	Length 429;
	Best local similarity	51.3%;	Pred. No. 0.82;		
	Matches	Conservative	0;	Mismatches	74;
				Indels	0;
				Gaps	0;
QY	24	TC TGGCCTCAGCTCGCGCGCCATGGACGCTGCCGAGGTGGAATCTCTGCCGAGAGG	83		
Db	429	TCAGCCTGGTGCCTGGAGATACCTCTACGGGGCGCTACTGGGATGCTCGCCGAGTTCG	370		
QY	84	AGCTGGTTACCATTTATCCCCAACTCTAGTCTGGACAAGATCTACCTCATCGGGGGGAC	143		
Db	369	ACGGGCTGCACCTTCGAGACCTGCTTCTACAGGGCCTTGGAAATCGCCATCGGCGCCGCC	310		
QY	144	TGGGGCCTTTTAAACCTCGTGGTTTACCGGTGAA	175		
Db	309	TGCGCGGCTTCGACGCGGGCGCCAGGGCGAA	278		

RESULT 14  
US-09-252-991A-13138  
; Sequence 13138, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13138  
; LENGTH: 711  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13138

```

Query Match      11.0%; Score 33; DB 4; Length 711;
Best Local Similarity 52.6%; Pred. No. 1.5;
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY      11  CTGTCGGTGAGGCTCTGGCTGCAGCTCGCGCCGCCATGGACGCTGCCGAGGTCGAATTC 70
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Db	136	CTGTCCGCGGTGCTGCTGGCTGGTGGCGGCACTGTACGCCACGGTAGTGCGCGGC	195
QY	71	CTCGCCGAGAGGAGCTGGTACATTTATCCCCAACTTCAGTCTGGCAAGAATCTACCTC	130
Db	196	ATCCCGGACCTGGTGCATCCTCTGATCTTCTACGGCGGCAGGGCCTGCTGAATGG	255
QY	131	ATCGGGGGGAGCCTGGG	147
Db	256	GTCGCCCGCAACTCGG	272

RESULT 15

```

US-09-252-991A-12804/G
; Sequence 12804, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12804
; LENGTH: 783
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12804

```

Query Match	11.0%	Score 33;	DB 4;	Length 783;
Best Local Similarity	52.6%;	Pred. No. 1.5;		
Matches	72;	Conservative	0;	Mismatches 65; Indels 0; Gaps 0;
QY	11	CTGTGGTGAGGCTCTGGCCTGCAGCTCGCGCGCCATGGACGCTGCCGAGGTGCAATTC	70	
Db	633	CTGTGCGCGTGCGCTGGCTGGCGCTGCTGGCGGACCTGTACGCCACGGTAGTCGCGCGC	574	
QY	71	CTCGCGGAGAAGGAGCTGTGTACCATTTATCCCAACTTCAGTCTGGACAAAGATCTACCTC	130	
Db	573	ATCCCGACCTGTGCTGATCCTGCTCATCTTTACGCGCGCCAGGSCCTGCTGAATGG	514	
QY	131	ATCGGGGGGACCTGGG	147	
Db	513	GTCCGCCCGCAACTCGG	497	

Search completed: October 14, 2003, 17:04:28  
Job time : 43.6667 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 14:17:11 ; Search time 143.333 Seconds  
(without alignments)  
5431.000 Million cell updates/sec

Title: US-09-854-124-6

Perfect score: 300

Sequence: 1 aattccgtgtcgtcgtga.....cgaagggaagaactttac 300.

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	300	100.0	300	9	US-09-854-124-6
2	284.8	94.9	597	10	US-09-796-692-4407
3	284.8	94.9	597	14	US-10-040-862-4407
4	284.8	94.9	1210	9	US-09-925-301-214
5	126.2	42.1	484	11	US-09-918-995-19938
6	102.6	34.2	178	10	US-09-783-590-12142
7	63	21.0	382	14	US-09-796-692-9460
8	63	21.0	382	14	US-10-040-862-9460
9	36.4	12.1	3008	12	US-10-205-219-83
10	35.4	11.8	582	14	US-10-156-761-2179
11	35.4	11.8	9025608	14	US-10-156-761-1
12	34.8	11.6	1018	10	US-09-896-856-6
13	34.8	11.6	1018	14	US-10-107-931-6
14	34.8	11.6	1539	10	US-09-901-540-1
15	34.8	11.6	1539	10	US-09-901-540-2
16	34.8	11.6	1539	10	US-09-901-257-1

c 17	34.8	11.6	1539	10	US-09-901-257-2	Sequence 2, Appli
18	34.8	11.6	1539	10	US-09-954-531-967	Sequence 967, App
19	34.8	11.6	3161	12	US-10-058-597-14	Sequence 14, Appl
c 20	34	11.3	501	14	US-10-156-761-7483	Sequence 7483, Ap
c 21	34	11.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
22	33.6	11.2	32846	12	US-10-017-161-1557	Sequence 1557, Ap
23	33.4	11.1	1118	10	US-09-974-300-2728	Sequence 2728, Ap
c 24	33	11.0	2000	10	US-09-938-842A-5378	Sequence 135, App
c 25	32.6	10.9	480	9	US-09-073-009-135	Sequence 135, App
c 26	32.6	10.9	480	9	US-09-793-306-135	Sequence 6871, Ap
27	32.6	10.9	1695	14	US-10-156-761-6871	Sequence 12, Appli
28	32.6	10.9	1709	12	US-10-174-133-12	Sequence 1244, Ap
29	32.4	10.8	885	14	US-10-156-761-3244	Sequence 649, App
30	32.2	10.7	45191	15	US-10-080-170-649	Sequence 2880, Ap
31	32	10.7	5721	14	US-10-156-761-2880	Sequence 15102, A
c 32	32	10.7	125746	14	US-10-156-761-15102	Sequence 631, App
33	31.8	10.6	534	14	US-10-156-761-631	Sequence 41006, A
34	31.8	10.6	540	13	US-10-027-632-41006	Sequence 74601, A
35	31.8	10.6	546	13	US-10-027-632-74601	Sequence 299544,
36	31.8	10.6	546	13	US-10-027-632-299544	Sequence 85, Appl
37	31.8	10.6	546	13	US-10-027-632-299543	Sequence 12, Appl
38	31.8	10.6	978	12	US-09-971-392-85	Sequence 33, Appl
39	31.8	10.6	1629	10	US-09-880-107-3799	Sequence 25, Appl
40	31.8	10.6	1629	10	US-09-880-107-3799	Sequence 1201, Ap
41	31.8	10.6	1829	14	US-10-146-473-12	Sequence 1917, Ap
42	31.8	10.6	10311	12	US-10-132-134-33	
43	31.8	10.6	50543	12	US-10-132-134-25	
44	31.6	10.5	379	10	US-09-880-107-1201	
45	31.6	10.5	930	14	US-10-156-761-1917	

#### ALIGNMENTS

#### RESULT 1

US-09-854-124-6  
; Sequence 6, Application US/09854124  
; Patent No. US20020076735A1

GENERAL INFORMATION:  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Escobedo, Jaime  
; APPLICANT: Innis, Michael A.  
; APPLICANT: Garcia, Pablo Dominquez  
; APPLICANT: Sudduth-Klinger, Julie  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Giese, Klaus  
; APPLICANT: Randazzo, Filippo  
; APPLICANT: Kennedy, Giulia C.  
; APPLICANT: Pot, David  
; APPLICANT: Kassam, Altaf  
; APPLICANT: Lamson, George  
; APPLICANT: Drmanac, Radoje  
; APPLICANT: Crkvenjakov, Radomir  
; APPLICANT: Dickson, Mark  
; APPLICANT: Drmanac, Shezana  
; APPLICANT: Labat, Ivan  
; APPLICANT: Leshkowitz, Dena  
; APPLICANT: Kita, David  
; APPLICANT: Garcia, Veronica  
; APPLICANT: Jones, Lee William  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Stache-Crain, Birgit  
; TITLE OF INVENTION: Diagnostic and Therapeutic Methods Using  
; TITLE OF INVENTION: Molecules Differentially Expressed in Cancer Cells  
; FILE REFERENCE: 2300-1490  
; CURRENT APPLICATION NUMBER: US/09/854,124  
; CURRENT FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: 09/400,947  
; PRIOR FILING DATE: 1999-09-22  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 300  
; TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-854-124-6

Query Match      100.0%; Score 300; DB 9; Length 300;
Best Local Similarity 100.0%; Pred. No. 8.6e-90;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTCGGTTCCTCGGTGAGGCTCTGGCTGCAGCTCGCGCGCCATGGAGCGTGCCTG 60
Db 1 AATTCGGTTCCTCGGTGAGGCTCTGGCTGCAGCTCGCGCGCCATGGAGCGTGCCTG 60
QY 61 GGTCAAAATTCCTCGCGGAGAGAGCTGTTACCAATATCCCAACTTCAGTCTGACAA 120
Db 61 GGTCAAAATTCCTCGCGGAGAGAGCTGTTACCAATATCCCAACTTCAGTCTGACAA 120
QY 121 GATCTACCTCATCGCGGGGGGACCTGGGGCCCTTTTAAACCTCGTGGAAAGTGC 180
Db 121 GATCTACCTCATCGCGGGGGGACCTGGGGCCCTTTTAAACCTCGTGGAAAGTGC 180
QY 181 CCTGTGGCTGGGATTAACCTGAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTG 240
Db 181 CCTGTGGCTGGGATTAACCTGAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTG 240
QY 241 GATGATCATAGAAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAGAACTTTTAC 300
Db 241 GATGATCATAGAAAGTTGGAGAGATGAGGATCATGAACGAAAGGAAGAACTTTTAC 300

RESULT 2
US-09-796-692-4407
; Sequence 4407, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT FILING DATE: 2001-03-01
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4407
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure

; LOCATION: (583)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (592)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-4407

Query Match      94.9%; Score 284.8; DB 10; Length 597;
Best Local Similarity 99.3%; Pred. No. 1.3e-84;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 GTCGGTAGGCTCTGGCTGCAGCTCGCGCGCCATGGAGCGTGCAGAGTGCATTCCT 72
Db 37 GACCGTAGGCTCTGGCTGCAGCTCGCGCGCCATGGAGCGTGCAGAGTGCATTCCT 96
QY 73 CGCCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAGATCTACCTCAT 132
Db 97 CGCCGAGAAGGAGCTGGTTACCATTTATCCCAACTTCAGTCTGGACAGATCTACCTCAT 156
QY 133 CGGGGGGACCTGGGGCCCTTTTAAACCTGGTTTACCCCTGGAAAGTGCCTGTGGCTGGC 192
Db 157 CGGGGGGACCTGGGGCCCTTTTAAACCTGGTTTACCCCTGGAAAGTGCCTGTGGCTGGC 216
QY 193 GATTAACTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGGATGGATGA 252
Db 217 GATTAACTGAAACAAAGACAGAAATGTGCGCTGCTCCCTCCAGAGTGGATGGATGA 276
QY 253 AAAGTTGGAGAAGATGAGGATCATGAACGAAAGGAAGAACTTTTAC 300
Db 277 AAAGTTGGAGAAGATGAGGATCATGAACGAAAGGAAGAACTTTTAC 324

RESULT 3
US-10-040-862-4407
; Sequence 4407, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT FILING DATE: 2001-11-06
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
```



```
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4407
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (583)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (592)
; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-4407

Query Match          94.9%   Score 284.8; DB 14; Length 597;
Best Local Similarity 99.3%; Pred. No. 1.7e-84;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      13  GTCCGCTAGGCTCTGGCCTCGACGTGCGCGGCCCATGGACGCTGCCGAGGTTCGAATTTCCT 72
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      37  GACCCTGAGCTCTGGCCTCGACTGCGCGGCCATGGAGCTGCCGAGTTCGAAATTCCT 96

Qy      73  CGCCGAGAAGAGCTGTTTACCATTATCCCACTTCAGTCTGGACAAGATCACCTCAT 132
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      97  CGCCGAGAAGAGCTGTTTACCATTATCCCACTTCAGTCTGGACAAGATCACCTCAT 156

Qy     133  CCGGGGGGACCTGGGGCTTTTAACCTTGTTTACCCGTGGAAGTGCCTGTGGCTGGC 192
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     157  CCGGGGGGACCTGGGGCTTTTAACCTTGTTTACCCGTGGAAGTGCCTGTGGCTGGC 216

Qy     193  GATTAACCTGAAAACAAGACAGAAATGTCGCCCTGCTCCCTCCAGAGTGGATGATGATA 252
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     217  GATTAACCTGAACAAGACAGAAATGTCGCCCTGCTCCCTCCAGAGTGGATGATGATA 276

Qy     253  AAGTTGGAGAAGATGAGGATCATGAACGAAAGAGAAACTTTTAC 300
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     277  AAGTTGGAGAAGATGAGGATCATGAACGAAAGAGAAACTTTTAC 324

RESULT 4
US-09-925-301-214
; Sequence 214, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 214
; LENGTH: 1210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-214

Query Match          94.9%   Score 284.8; DB 9; Length 1210;
Best Local Similarity 99.3%; Pred. No. 1.7e-84;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      13  GTCCGCTAGGCTCTGGCCTCGACGTGCGCGGCCCATGGACGCTGCCGAGGTTCGAATTTCCT 72
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      47  GACCCTGAGCTCTGGCCTCGACTGCGCGGCCCATGGACGCTGCCGAGGTTCGAATTTCCT 106

Qy      73  CGCCGAGAAGAGCTGTTTACCATTATCCCACTTCAGTCTGGACAAGATCACCTCAT 132
```

107 CGCGAGAGAGAGCTGGTTACCATATATCCCCAACCTTCAGTCTGGACAAGATCTACCTCAT 167

QY 133 CGGGGGGACCTGGGGCCCTTTTAAACCCCTGGTTTACCCCTGGAAGTGCCTCTGGCTGCGC 192

Db 167 CGGGGGGACCTGGGGCCCTTTTAAACCCCTGGTTTACCCCTGGAAGTGCCTCTGGCTGCGC 226

QY 193 GATTAACTGAAACAAGACAGAAATGTGCCTGTCTCCAGAGTGGATGGATGTAGA 252

Db 227 GATTAACTGAAACAAGACAGAAATGTGCCTGTCTCCAGAGTGGATGGATGTAGA 286

QY 253 AAGTTGGAGAGATGAGGGATCATGAACGAAGGAAGAACTTTTAC 300

Db 287 AAGTTGGAGAGATGAGGGATCATGAACGAAGGAAGAACTTTTAC 334

RESULT 5

US-09-918-995-19938

; Sequence 19938, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 19938

; LENGTH: 484

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-918-995-19938

Query Match 42.1%; Score 126.2; DB 11; Length 484;

Best Local Similarity 97.7%; Pred. No. 8.6e-32;

Matches 128; Conservative 0; Mismatches 3; Indels 0; Gaps 0

QY 170 GTGGAAGTGCCTGTGCTGGCGATTACCTGAAACAAGACAGAAATGTGCCTGCTC 229

Db 37 CGGGAAGTGCCTGTGCTGGCGATTACCTGAAACAAGACAGAAATGTGCCTGCTC 96

QY 230 CTTCCAGATGGATGTAGAAAAGTTGGAGAAGATGAGGATCATCAACGAAGGAA 289

Db 97 CTTCCAGATGGATGTAGAAAAGTTGGAGAAGATGAGGATCATCAACGAAGGAA 156

QY 290 GAAACTTTTAC 300

Db 157 GAAACTTTTAC 167

RESULT 6

US-09-783-590-12142

; Sequence 12142, Application US/09783590

; Patent No. US20020110850A1

; GENERAL INFORMATION:

; APPLICANT: Dillion, Patrick J.

; APPLICANT: Haseitine, William A.

; APPLICANT: Li, Haodong

; APPLICANT: Rosen, Craig A.

; APPLICANT: Ruben, Steven M.

; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2

; FILE REFERENCE: PO-16.2C1

; CURRENT APPLICATION NUMBER: US/09/783,590

; CURRENT FILING DATE: 2000-02-15

; PRIOR APPLICATION NUMBER: 08/420,856

; PRIOR FILING DATE: 1995-04-12

; PRIOR APPLICATION NUMBER: 08/346,731

; PRIOR FILING DATE: 1994-11-21

; NUMBER OF SEQ ID NOS: 12485

; SOFTWARE: Patent In Ver. 2.0

```

; SEQ ID NO 12142
; LENGTH: 178
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (49)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (61)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (66)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (143)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (166)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (171)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-12142

Query Match          34.2%; Score 102.6; DB 10; Length 178;
Best Local Similarity 80.2%; Pred. No. 4.2e-24;
Matches 142; Conservative 0; Mismatches 30; Indels 5; Gaps 2;

QY 47 ATGGACGCTGCGAGGTGCAATTCCTCGCGAGAGGAGCTGGTTACCATTTATCCCCAAC 106
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 ATGAACGCTGCGAAGTCAATTCCTCGCGAAGAGGAGCTGGTTACNATTATCCCCAAN 61

QY 107 TTCATCTGCGACAAGATCTACCTCATCGGGGGGGA--CCTGGGGCCTTTTAAACCTGGTT 164
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 TTAATTTTGGCAAAATCTACCTATATGGGGGGGAACCTGGGGCCTTTTAAACCTGGTT 121

QY 165 TACCGGTGAA---GTGCCCTGTGGTGGCGATTAACTGAACAAAGACAGAAAT 218
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 TACCGGTGGAGTGCCCTTGTGGTGGGAATTAACCTGAANAAGNCAGAAAT 178

RESULT 7
US-09-796-692-9460/c
; Sequence 9460, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796.692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
;
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9460
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (72)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (80)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (89)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (95)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (122)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (129)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (131)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (155)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (173)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (222)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (235)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (236)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (237)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (245)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (251)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (261)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (305)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (312)
; OTHER INFORMATION: n=A,T,C or G

```

```
; NAME/KEY: unsure
; LOCATION: (321)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (334)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (352)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (354)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (373)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (375)
; OTHER INFORMATION: n=A,T,C or G
; US-09-796-692-9460

Query Match      21.0%; Score 63; DB 10; Length 382;
Best Local Similarity 91.3%; Pred. No. 9e-11;
Matches 63; Conservative 0; Mismatches 6; Indels 0; Caps 0;

Qy 232 TCCAGAGTGGATGATGAGAAAGTTGGAGAGATGAGGGATCATGACGAAAGGAAGA 291
Db 382 TCCAGAGNATGGATGATGAGAAAGTTGNANAAGATGAGGGATCATGACGAAAGGAAGA 323
Qy 292 AACTTTATAC 300
Db 322 ANCTTTATAC 314

RESULT 8
US-10-040-862-9460/c
; Sequence 9460, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; NAME/KEY: unsure
; LOCATION: (321)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (334)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (352)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (354)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (373)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (375)
; OTHER INFORMATION: n=A,T,C or G
; US-09-796-692-9460
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9460
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (72)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (80)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (89)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (95)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (122)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (129)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (131)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (155)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (173)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (222)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (235)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (236)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (237)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (245)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (251)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
```

```

; LOCATION: (261)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (305)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (312)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (321)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (334)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (352)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (354)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (373)
; OTHER INFORMATION: n-A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (375)
; OTHER INFORMATION: n-A,T,C or G
; US-10-040-862-9460

```

```

Query Match      21.0%; Score 63; DB 14; Length 382;
Best Local Similarity 91.3%; Pred. No. 9e-11;
Matches 63; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 232 TCCAGAGTGGATGATGAGAAAGTTGGAGAAGATGAGGATCATGATGACCAAGGAGA 291
Db 382 TCCAGAGTGGATGATGAGAAAGTTGNANAGATGAGGATCATGATGACCAAGGAGA 323
QY 292 AACTTTTAC 300
Db 322 ANCTTTTAC 314

```

```

RESULT 9
US-10-205-219-83/c
; Sequence 83, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pincock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205.219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 83
; LENGTH: 3008
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: N-G, N-G dimethylarginine dimethylaminohydrolase

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2411, 2412
; OTHER INFORMATION: n is a or g or c or t
; US-10-205-219-83

Query Match      12.1%; Score 36.4; DB 12; Length 3008;
Best Local Similarity 46.2%; Pred. No. 0.15;
Matches 121; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 17 GTGAGGCTCTGGCTCGAGCTCGCGCCGCCATGAGACGCTCGGAGGTGCAATTCCTCGCC 76
Db 740 GTCAACCTCTCTCTCGTGGGCTAGGCGCCCGGGGGGTGATGAGGCGGTCTCTCTCGCA 681
QY 77 GAGAAGGAGCTGGTTACCAATTATCCCAACTTTCAGTCTGGACAAGATCTACCTCATCGGG 136
Db 680 CACCACGGCCACGTCCTCCACGACACGCACTCAGGAGGCTCTCGTGGCGCCGACGCTG 621
QY 137 GGGACCTGGGGCCTTTTAAACCCCTGGTGTACCCCTGGAAAGTCCCTCTGTGCTGGCGATT 196
Db 620 CACCACCTGACGCCCCAGCTTGTCTGCCAGCACGCCACGCTAGAGCTGGTCTGGCGCTC 561
QY 197 AACCTGAAACAAGACAGAAATGCGCTGCTCTCCCTCCAGAGTGGATGGATGTAGAAAAG 256
Db 560 AGCGGAGCGAAATCCACCTCTTCGCCCTGGGAGCGCTCAGCGGTGGCGGCACAGGGA 501
QY 257 TTGGAGAAGATGAGGATCATG 278
Db 500 CTCGGCGGAGCCCGCACCAG 479

```

## RESULT 10

```

US-10-156-761-2179
; Sequence 2179, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156.761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2179
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(582)
; US-10-156-761-2179

```

```

Query Match      11.8%; Score 35.4; DB 14; Length 582;
Best Local Similarity 63.5%; Pred. No. 0.17;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 40 CGCGCCATGAGCGCTCCGAGTTCGAATTCCTCGCGAGAAGAGCTGGTTACCATAT 99
Db 159 CGCGGCTGGACGCTCGACGACCTCGAGCGATCGCGAGAAGAGCCACCGCTGGC 218
QY 100 CCCCACCTTCAGTCTGACAGATC 124
Db 219 CCTCGACTTCGCGGGATCAGATC 243

```

```
RESULT 11
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      11.8%; Score 35.4; DB 14; Length 9025608;
Best Local Similarity 53.5%; Pred. No. 7.6;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 40 CGCGCCATGGAGCTGCGGAGTGAATTCCTCGCGAGAGGAGCTGGTTACCATTTAT 99
Db 2665457 CGCGCGCCCTGGAGCTGAGAGCTCGAGCATCGCGGAGAGAGCGCACCGCTGCG 2665516

QY 100 CCCCACCTTCAGTCTGGACAAGATC 124
Db 2665517 CCTCGACTTCGCGGATCACGATC 2665541

RESULT 12
US-09-896-856-6
; Sequence 6, Application US/09896856
; Patent No. US20020137189A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; Chien, Kenneth
; King, Kathleen
; Pennica, Diane
; Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/896,856
; FILING DATE: 29-Jun-2001
; CLASSIFICATION: <Unknown>
```

```
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/733,850
FILING DATE: 18-OCT-1996
APPLICATION NUMBER: US 08/471,112
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: 08/233,609
FILING DATE: 25-APR-1994
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
APPLICATION NUMBER: 08/443129
FILING DATE: 17-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P0894PID2C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-2066
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-896-856-6

Query Match      11.6%; Score 34.8; DB 10; Length 1018;
Best Local Similarity 50.6%; Pred. No. 0.33;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 21 GGCTCTGGCTGCAGCTCGCGCCATGCGCATGAGGCTCGCGAGTTCCTCGCGGAGA 80
Db 484 GGGCGAGCCCGCCCGCCAGCGCTCAGCGCGCTCCGCCACCGGGGTCTTCCCGCCA 543

QY 81 AGAGCTGTTTACCATTTATCCCAACTTCAGTCTGGACAAGATCTACCTCATCGGGGGG 140
Db 544 AGGTGTGGGGTCTCGCGTTCGGGCTCTACCGGAGTGGGTGAGCCCGCAGCGGCGG 603

QY 141 ACCTGGGGCTTTTAACTCCCTGGTTTACCCTGGAAGTGCCTGTG 186
Db 604 ACCTGGGCGAGCTGCTGCGCGGGGCTCGCGCTGAGCCCGCGGGG 649

RESULT 13
US-10-107-931-6
; Sequence 6, Application US/10107931
; Publication No. US20030054550A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; Chien, Kenneth
; King, Kathleen
; Pennica, Diane
; Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/107,931
; FILING DATE: 26-Mar-2002
; CLASSIFICATION: <Unknown>
```

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/364,187  
; FILING DATE: 30-Jul-1999  
; APPLICATION NUMBER: 08/286,304  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hasak, Janet E.  
; REGISTRATION NUMBER: 28,616  
; REFERENCE/DOCKET NUMBER: 894P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1896  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1018 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-107-931-6

Query Match 11.6%; Score 34.8; DB 14; Length 1018;  
Best Local Similarity 50.6%; Pred. No. 0.33;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 21 GGCTCTGGCCTGCAGCTCGCGCCGCAATGAGCGTGGCGAGGTGGAATTCCTCGCCGAGA 80  
DB 484 GGGCGAGCGCCCGCCGCGCCGCTCAGCGGCTCGGCCACCGGGGTCTTCCCGCCA 543  
QY 81 AGGAGCTGGTTACCAATTATCCCAACTTCACTCTGGACAAGATCTACCTCATCGGGGGG 140  
DB 544 AGGTGCTGGGCTCGCGCTTTCGCGGCTCTACCGGAGTGGCTGAGCGCGCAGGAGGCG 603  
QY 141 ACCTGGGGCCCTTTAAACCTGTTTACCCGTTGGAAGTGCCCTGTG 186  
DB 604 ACCTGGGCGAGCTGCTGCCGGGGGCTCGGCTGAGCGCGCGGGG 649

RESULT 14  
US-09-901-540-1  
; Sequence 1, Application US/09901540  
; Patent No. US20020102622A1  
; GENERAL INFORMATION:  
; APPLICANT: Botstein, David  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Lawrence, David A.  
; APPLICANT: Pennica, Diane  
; APPLICANT: Roy, Margaret A.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: Cardiotrophin-1 Compositions and Methods for the  
; FILE REFERENCE: P2533D2  
; CURRENT APPLICATION NUMBER: US/09/901,540  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 09/648,258  
; PRIOR FILING DATE: 2000-08-25  
; PRIOR APPLICATION NUMBER: US 09/033,114  
; PRIOR FILING DATE: 1998-03-02  
; PRIOR APPLICATION NUMBER: US 08/733,850  
; PRIOR FILING DATE: 1996-10-18  
; PRIOR APPLICATION NUMBER: US 08/443,129  
; PRIOR FILING DATE: 1995-05-17  
; PRIOR APPLICATION NUMBER: US 08/286,304  
; PRIOR FILING DATE: 1994-08-05  
; PRIOR APPLICATION NUMBER: US 08/233,609  
; PRIOR FILING DATE: 1994-04-25  
; PRIOR APPLICATION NUMBER: US 60/113,296  
; PRIOR FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 1  
; LENGTH: 1539  
; TYPE: DNA  
; ORGANISM: Human  
US-09-901-540-2

; ORGANISM: Human  
US-09-901-540-1  
Query Match 11.6%; Score 34.8; DB 10; Length 1539;  
Best Local Similarity 50.6%; Pred. No. 0.39;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 21 GGCTCTGGCCTGCAGCTCGCGCCGCAATGAGCGTGGCGAGGTGGAATTCCTCGCCGAGA 80  
DB 484 GGGCGAGCGCCCGCCGCGCCGCTCAGCGGCTCGGCCACCGGGGTCTTCCCGCCA 543  
QY 81 AGGAGCTGGTTACCAATTATCCCAACTTCACTCTGGACAAGATCTACCTCATCGGGGGG 140  
DB 544 AGGTGCTGGGCTCGCGCTTTCGCGGCTCTACCGGAGTGGCTGAGCGCGCAGGAGGCG 603  
QY 141 ACCTGGGGCCCTTTAAACCTGTTTACCCGTTGGAAGTGCCCTGTG 186  
DB 604 ACCTGGGCGAGCTGCTGCCGGGGGCTCGGCTGAGCGCGCGGGG 649

RESULT 15  
US-09-901-540-2/c  
; Sequence 2, Application US/09901540  
; Patent No. US20020102622A1  
; GENERAL INFORMATION:  
; APPLICANT: Botstein, David  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Lawrence, David A.  
; APPLICANT: Pennica, Diane  
; APPLICANT: Roy, Margaret A.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: Cardiotrophin-1 Compositions and Methods for the  
; FILE REFERENCE: P2533D2  
; CURRENT APPLICATION NUMBER: US/09/901,540  
; CURRENT FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 09/648,258  
; PRIOR FILING DATE: 2000-08-25  
; PRIOR APPLICATION NUMBER: US 09/033,114  
; PRIOR FILING DATE: 1998-03-02  
; PRIOR APPLICATION NUMBER: US 08/733,850  
; PRIOR FILING DATE: 1996-10-18  
; PRIOR APPLICATION NUMBER: US 08/443,129  
; PRIOR FILING DATE: 1995-05-17  
; PRIOR APPLICATION NUMBER: US 08/286,304  
; PRIOR FILING DATE: 1994-08-05  
; PRIOR APPLICATION NUMBER: US 08/233,609  
; PRIOR FILING DATE: 1994-04-25  
; PRIOR APPLICATION NUMBER: US 60/113,296  
; PRIOR FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 2  
; LENGTH: 1539  
; TYPE: DNA  
; ORGANISM: Human  
US-09-901-540-2

Query Match 11.6%; Score 34.8; DB 10; Length 1539;  
Best Local Similarity 50.6%; Pred. No. 0.39;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 21 GGCTCTGGCCTGCAGCTCGCGCCGCAATGAGCGTGGCGAGGTGGAATTCCTCGCCGAGA 80  
DB 1056 GGGCGAGCGCCCGCCGCGCCGCTCAGCGGCTCGGCCACCGGGGTCTTCCCGCCA 997  
QY 81 AGGAGCTGGTTACCAATTATCCCAACTTCACTCTGGACAAGATCTACCTCATCGGGGGG 140  
DB 996 AGGTGCTGGGCTCGCGCTTTCGCGGCTCTACCGGAGTGGCTGAGCGCGCAGGAGGCG 937  
QY 141 ACCTGGGGCCCTTTAAACCTGTTTACCCGTTGGAAGTGCCCTGTG 186  
DB 936 ACCTGGGCGAGCTGCTGCCGGGGGCTCGGCTGAGCGCGCGGGG 891

Search completed: October 14, 2003, 17:17:44  
Job time : 154.333 secs

---

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 13:00:56 ; Search time 1477 Seconds  
(without alignments)  
8309.335 Million cell updates/sec

Title: US-09-854-124-5  
Perfect score: 300  
Sequence: 1 acgaatccgagccctggc.....tgaggtaactcgtggtctgg 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenBank:

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.ma.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htg\_mus.\*

34: em.htg\_pln.\*

35: em.htg\_rod.\*

36: em.htg\_nam.\*

37: em.htg\_vrt.\*

38: em.sy.\*

39: em.htgo\_hum.\*

40: em.htgo\_mus.\*

41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	300	100.0	858	6	BD156103	BD156103 Primer fo
2	300	100.0	858	9	AK001275	AK001275 Homo sapi
3	300	100.0	1174	9	AF151880	AF151880 Homo sapi
4	300	100.0	1180	9	BC010164	BC010164 Homo sapi
5	300	100.0	1188	9	AF201939	AF201939 Homo sapi
6	298.4	99.5	1203	9	BC003186	BC003186 Homo sapi
7	292.4	97.5	667	6	BD147616	BD147616 Primer fo
8	258.4	86.1	116792	9	HS1100H13	AL035419 Human DNA
9	201	67.0	160356	9	AC009068	AC009068 Homo sapi
10	201	67.0	185321	9	AC123908	AC123908 Homo sapi
11	201	67.0	245210	2	AC137771	AC137771 Homo sapi
12	199.4	66.5	3097	9	AK091519	AK091519 Homo sapi
13	140	46.7	454	6	AX336730	AX336730 Sequence
14	125.8	41.9	933	5	AB097169	AB097169 Xenopus l
15	111.8	37.3	207307	2	AC117085	AC117085 Rattus no
16	104.4	34.8	188292	2	AC103360	AC103360 Mus muscu
17	66	22.0	362	6	AX396467	AX396467 Sequence
18	53.8	17.9	825	3	AY071234	AY071234 Drosophill
19	53.8	17.9	72722	3	AC004373	AC004373 Drosophill
20	53.8	17.9	132666	3	AC005270	AC005270 Drosophill
21	53.8	17.9	182897	3	AC092232	AC092232 Drosophill
22	53.8	17.9	218794	2	AC018064	AC018064 Drosophill
23	53.8	17.9	292393	3	AE003576	AE003576 Drosophill
24	44.6	14.9	165158	2	AC141687	AC141687 Apis mell
25	41.6	13.9	213	9	AF534827	AF534827 Homo sapi
26	41.6	13.9	92814	9	AL391903	AL391903 Human DNA
27	41.6	13.9	210842	2	AC140748	AC140748 Rattus no
28	41.6	13.9	219443	2	AC023320	AC023320 Homo sapi
29	41.6	13.9	246542	2	AC125601	AC125601 Rattus no
30	39.6	13.2	734	6	BD150236	BD150236 Primer fo
31	39.6	13.2	1150	9	AF534824	AF534824 Homo sapi
32	39.6	13.2	3071	6	BD158620	BD158620 Primer fo
33	39.6	13.2	3071	9	AK023755	AK023755 Homo sapi
34	39.6	13.2	91511	9	AL133404	AL133404 Human DNA
35	39.6	13.2	92814	9	AL391903	AL391903 Human DNA
36	36.2	12.1	162641	2	AC091406	AC091406 Rattus no
37	36	12.0	654	6	AX386100	AX386100 Sequence
38	36	12.0	49995	2	AC011335	AC011335 Homo sapi
39	36	12.0	77139	2	AC008466	AC008466 Homo sapi
40	36	12.0	120849	2	AC087148	AC087148 Mus muscu
41	36	12.0	128928	9	AC068658	AC068658 Homo sapi
42	36	12.0	157208	9	AC113409	AC113409 Homo sapi
43	36	12.0	203418	2	AC063958	AC063958 Homo sapi
44	35.6	11.9	1129	8	AF499724	AF499724 Thellungi
45	35.6	11.9	215707	2	AC099096	AC099096 Rattus no

ALIGNMENTS

RESULT 1  
BD156103

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BD156103  
Primer for synthesizing full-length cDNA and use thereof.  
BD156103  
BD156103.1 GI:27861861  
JP 2002191363-A/10946.  
Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 858)

Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,

Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.

Primer for synthesizing full-length cDNA and use thereof

858 bp DNA linear RNT 17-JAN-2003

JOURNAL Patent: JP 2002191363-A 10946 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 COMMENT OS Homo sapiens (human)  
 PN JP 2002191363-A/10946  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOgai, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC  
 10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 PC Primer for synthesizing full-length cDNA and use thereof FH key

FT CDS Location/Qualifiers  
 1..858  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 227 a 214 c 222 g 195 t

BASE COUNT 227 a 214 c 222 g 195 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 6; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-69;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGACCTGTCAAGGATATGTGGGACATCGTATAGCCAACTCCGAGTGT 60  
 DB 368 ACGAAATCCGACCTGTCAAGGATATGTGGGACATCGTATAGCCAACTCCGAGTGT 427

QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 120  
 DB 428 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 487

QY 121 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACCATGTACAACTCC 180  
 DB 488 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACCATGTACAACTCC 547

QY 181 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
 DB 548 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 607

QY 241 GCAGCGGCTTGTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGTGTTCTGG 300  
 DB 608 GCAGCGGCTTGTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGTGTTCTGG 667

RESULT 2  
 AK001275 858 bp mRNA linear PRI 01-AUG-2002  
 LOCUS Homo sapiens cDNA FLJ10413 fis, clone NT2RP1000063.  
 DEFINITION  
 ACCESSION AK001275  
 VERSION AK001275.1 GI:7022428  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
 ISOGAI, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
 Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,  
 Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,  
 Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,  
 Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,  
 Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,  
 Nakamura, Y., Nagahari, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 858)  
 ISOGAI, T. and OTSUKI, T.

TITLE  
 JOURNAL  
 REFERENCE 2  
 AUTHORS

Direct Submission  
 Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 297-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection;  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 FEATURES Location/Qualifiers  
 1..858  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP1000063"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP1"  
 /note="cloning vector: pUC19FL3-mRNA from NT2 neuronal  
 precursor cells after 48-hours retinoic acid (RA)  
 induction."  
 37..594  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAA91595.1"  
 /db\_xref="GI:7022429"  
 /translation="MDAAVEFLAEKELVLIIPNFDLKIYLGIGDLPNPGPLVEV  
 PLMLAIKQKQKRLPPEMDVDEKLEMRHERKETFTPMPSYMETLKLHLNH  
 ASDNIPKAEIRTLVDMWDTIRIAKRVSDSFVROQFAHAKLDNLTIMEINTSGTFL  
 TQALNNHMYKLRNLQPLETSQSDP"  
 BASE COUNT 227 a 214 c 222 g 195 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-69;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGACCTGTGTCAAGGATATGTGGGACATCGTATAGCCAACTCCGAGTGT 60  
 DB 368 ACGAAATCCGACCTGTGTCAAGGATATGTGGGACATCGTATAGCCAACTCCGAGTGT 427

QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 120  
 DB 428 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 487

QY 121 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACCATGTACAACTCC 180  
 DB 488 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACCATGTACAACTCC 547

QY 181 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
 DB 548 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 607

QY 241 GCAGCGGCTTGTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGTGTTCTGG 300  
 DB 608 GCAGCGGCTTGTCTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCTGTGTTCTGG 667

RESULT 3  
 AF151880 1174 bp mRNA linear PRI 18-MAY-2000  
 LOCUS Homo sapiens CGI-122 protein mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF151880  
 VERSION AF151880.1 GI:4929712  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1174)  
 LAI, C.H., Chou, C.Y., Ch'ang, L.Y., Liu, C.S. and Lin, W.

**TITLE** Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics

**JOURNAL MEDLINE** Genome Res. 10 (5), 703-713 (2000)

**PUBMED** 2072150

**REFERENCE** 10810093

**AUTHORS** Lin W.-C.

**TITLE** Direct Submission

**JOURNAL** Submitted (17-MAY-1999) Institute of Biomedical Sciences, Academia Sinica, No. 128, Sec. II, Academia Road, Taipei 115, Taiwan

**FEATURES** Location/Qualifiers

1..1174

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="16"

77..634

/codon\_start=1

/product="CGI-122 protein"

/protein\_id="AD34117.1"

/db\_xref="GI:4929713"

/translation="MDAAEVEFLAEKELVTIIPNFSLDKIYLLIGDGLGFPNPLPVEV PLWLAINIKQKOKKRLPPWMDVEKLMRDRHEKEETFTPMPSPYMYELTKLLLNH ASDNIPKADIEITLVKMDWTRIAKLRSADSFVRQEAHAKLDNLTLWEINTSGTFL TQALNHYKLRITNLPLESTQSDP"

**BASE COUNT** 300 a 299 c 307 g 268 t

**ORIGIN**

Query Match 100.0%; Score 300; DB 9; Length 1174;

Best Local Similarity 100.0%; Pred. No. 4.5e-69;

Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAATCGGACCTGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60

DB 408 ACGAATCGGACCTGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 467

QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATTAACCTTGA 120

DB 468 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATTAACCTTGA 527

QY 121 TGGAGATCAACACACAGCGGACTTTCCTCACACAGCGCTCAACACATGTACAACTCC 180

DB 528 TGGAGATCAACACACAGCGGACTTTCCTCACACAGCGCTCAACACATGTACAACTCC 587

QY 181 GCACGAACCTCAGGCTCTGGAGAGTACTCACTCTCAGGACTTCTAGAGAAAGGCTGT 240

DB 588 GCACGAACCTCAGGCTCTGGAGAGTACTCACTCTCAGGACTTCTAGAGAAAGGCTGT 647

QY 241 GCAGCGGCTTCTGGGATGTGAGCGCTCAGGAGTACTCGGTCTGCTGG 300

DB 648 GCAGCGGCTTCTGGGATGTGAGCGCTCAGGAGTACTCGGTCTGCTGG 707

**RESULT 4**

**LOCUS** BC010164 1180 bp mRNA linear PRI 12-JUL-2001

**DEFINITION** Homo sapiens, HSPC037 protein, clone MGC:19836 IMAGE:4098007, mRNA, complete cds.

**ACCESSION** BC010164

**VERSION** BC010164.1 GI:14603431

**KEYWORDS** MGC.

**SOURCE** Homo sapiens (human)

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1180)

Strausberg, R.

**REFERENCE** Direct Submission

**AUTHORS** Submitted (02-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK** NIH-MGC Project URL: <http://mgc.nci.nih.gov>

## COMMENT

Contact: MGC help desk  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 27 Row: 1 Column: 5  
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.

## FEATURES

Location/Qualifiers

1..1180

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="LocusID:51659"

/db\_xref="taxon:9606"

/clone="MGC:19836 IMAGE:4098007"

/tissue\_type="Brain, neuroblastoma"

/clone\_lib="NIH\_MGC\_19"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

68..625

/codon\_start=1

/product="HSPC037 protein"

/protein\_id="AAH10164.1"

/db\_xref="GI:14603432"

/translation="MDAAEVEFLAEKELVTIIPNFSLDKIYLLIGDGLGFPNPLPVEV PLWLAINIKQKOKKRLPPWMDVEKLMRDRHEKEETFTPMPSPYMYELTKLLLNH ASDNIPKADIEITLVKMDWTRIAKLRSADSFVRQEAHAKLDNLTLWEINTSGTFL TQALNHYKLRITNLPLESTQSDP"

**BASE COUNT** 316 a 292 c 307 g 265 t

**ORIGIN**

## CDS

Query Match 100.0%; Score 300; DB 9; Length 1180;

Best Local Similarity 100.0%; Pred. No. 4.5e-69;

Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAATCGGACCTGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60

DB 399 ACGAATCGGACCTGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 458

QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATTAACCTTGA 120

DB 459 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATTAACCTTGA 518

QY 121 TGGAGATCAACACACAGCGGACTTTCCTCACACAGCGCTCAACACATGTACAACTCC 180

DB 519 TGGAGATCAACACACAGCGGACTTTCCTCACACAGCGCTCAACACATGTACAACTCC 578

QY 181 GCACCAACCTCCAGCCTCTGGAGAGTACTCACTCTCAGGACTTCTAGAGAAAGGCTGT 240

DB 579 GCACCAACCTCCAGCCTCTGGAGAGTACTCACTCTCAGGACTTCTAGAGAAAGGCTGT 638

QY 241 GCAGCGGCTTCTGGGATGTGAGCGCTCAGGAGTACTCGGTCTGCTGG 300

DB 639 GCAGCGGCTTCTGGGATGTGAGCGCTCAGGAGTACTCGGTCTGCTGG 698

## RESULT 5

AF201939  
 LOCUS AF201939 1188 bp mRNA linear PRI 20-JUL-2000  
 DEFINITION Homo sapiens DC5 (DC5) mRNA, complete cds.  
 ACCESSION AF201939  
 VERSION AF201939.1 GI:9295181  
 KEYWORDS FLI\_CDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 1188)  
 Li.Y., Peng.Y., Li.N., Gu.W., Han.Z., Fu.G. and Chen.Z.  
 Novel genes expressed in human dentritic cell  
 Unpublished  
 2 (bases 1 to 1188)  
 Li.Y., Peng.Y., Li.N., Gu.W., Han.Z., Fu.G. and Chen.Z.  
 Direct Submission  
 Submitted (05-NOV-1999) Chinese National Human Genome Center at  
 Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,  
 Shanghai 201203, P. R. China  
 JOURNAL

FEATURES  
 Location/Qualifiers  
 1..1188  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cell\_type="dentritic cell"  
 1..1188  
 /gene="DC5"  
 77..634  
 /gene="DC5"  
 /codon\_start=1  
 /product="DC5"  
 /protein\_id="AAF86875.1"  
 /db\_xref="GI:9295182"  
 /translation="MDAAEVEFLAEKELVTIIPNFSLDKIYLGIDLPNPGIPVEV  
 PLWAINLKQKORCLLPPEMDVEKLEKMRDHERKEETTPMPSPYMBELTKLLNH  
 ASDNPKADEIRTLVDMWTRIAKLRSVDSFVRQOEAKLDNLTLMEINTSGTFL  
 TQALNHMYKLRTNLQPLETSQSDP"  
 BASE COUNT 314 a 299 c 307 g 268 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 1188;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-69;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 Db 408 ACGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 67

Qy 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGACACATGCCAAGCTGGATAACTTGACCTTGA 120  
 Db 468 CTGCTGACAGCTTTGTGAGACAGCAGGAGGACACATGCCAAGCTGGATAACTTGACCTTGA 527

Qy 121 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 180  
 Db 528 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 587

Qy 181 GCAGAACCTCCACCCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
 Db 588 GCAGAACCTCCACCCCTCTGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 647

Qy 241 GCAGGCGGCTTGCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTGG 300  
 Db 648 GCAGGCGGCTTGCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTGG 707

RESULT 6  
 BC003186  
 LOCUS BC003186 1203 bp mRNA linear PRI 12-JUL-2001  
 DEFINITION Homo sapiens, HSPC037 protein, clone MGC:673 IMAGE:3503268, mRNA,  
 complete cds.  
 ACCESSION BC003186  
 VERSION BC003186.1 GI:13112024

KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REMARK  
 COMMENT

MGC.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 1203)  
 Strausberg,R.  
 Direct Submission  
 Submitted (13-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Seedei, Jacqueline  
 Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAL Plate: 6 Row: d Column: 16  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4929712.

FEATURES  
 source

Location/Qualifiers  
 1..1203  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="LocustID:51659"  
 /db\_xref="taxon:9606"  
 /clone="MGC:673 IMAGE:3503268"  
 /tissue\_type="Brain, neuroblastoma"  
 /clone\_11b="NIH\_MGC\_19"  
 /lab\_host="DH10B-R"  
 /note="vector: pOTB7"  
 85..642  
 /codon\_start=1  
 /product="HSPC037 protein"  
 /protein\_id="AAH03186.1"  
 /db\_xref="GI:13112025"  
 /translation="MDAAEVEFLAEKELVTIIPNFSLDKIYLGIDLPNPGIPVEV  
 PLWAINLKQKORCLLPPEMDVEKLEKMRDHERKEETTPMPSPYMBELTKLLNH  
 ASDNPKADEIRTLVDMWTRIAKLRSVDSFVRQOEAKLDNLTLMEINTSGTFL  
 TQALNHMYKLRTNLQPLETSQSDP"  
 BASE COUNT 326 a 300 c 309 g 268 t  
 ORIGIN

Query Match 99.5%; Score 298.4; DB 9; Length 1203;  
 Best Local Similarity 99.7%; Pred. No. 1.2e-68;  
 Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ACGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 Db 416 ACGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 475

Qy 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGACACATGCCAAGCTGGATAACTTGACCTTGA 120  
 Db 476 CTGCTGACAGCTTTGTGAGACAGCAGGAGGACACATGCCAAGCTGGATAACTTGACCTTGA 535

Qy 121 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 180  
 Db 536 TGGAGATCAACACACAGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 595

QY	181	GCACGACCTCCAGCCTCTCGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT	240
Db	596	GCACAAACCTCCAGCCTCTCGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT	655
QY	241	GCAGGCGGCTTGTCTGGGGATGTGAGCGCTCAGGAGTATGATGAGTACTCGTGGTCTTCGG	300
Db	656	GCAGGCGGCTTGTCTGGGGATGTGAGCGCTCAGGAGTATGATGAGTACTCGTGGTCTTCGG	715
RESULT 7			
BD147616			
LOCUS	BD147616	667 bp	DNA linear PAT 17-JAN-2003
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.		
ACCESSION	BD147616		
VERSION	BD147616.1	GI:27853374	
KEYWORDS	JP 2002191363-A/2459.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.		
TITLE	Primer for synthesizing full-length cDNA and use thereof		
JOURNAL	Patent: JP 2002191363-A 2459 09-JUL-2002;		
COMMENT	HELIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002191363-A/2459 PD 09-JUL-2002 PF 28-JUL-2000 JP 2000280990 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO, PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, PI KEIICHI NAGAI, TETSUJI OTSUKI PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC 10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key		
FT	source	1..667	
FEATURES	Location/Qualifiers	/organism='Homo sapiens (human)'	
source	1..667		
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
BASE COUNT	171 a	176 c	177 g 137 t 6 others
ORIGIN			
Query Match	97.5%; Score 292.4; DB 6; Length 667;		
Best Local Similarity	97.7%; Pred. No. 5.2e-67;		
Matches 293;	Conservative	0;	Mismatches 7; Indels 0; Gaps 0;
QY	1	ACGAATCCGGACCTCTGGAGATATGTGGACACTCGTATAGCCAACTCCGAGTGT	60
Db	367	ACGAATCCGGACCTCTGGAGATATGTGGACACTCGTATAGCCAACTCCGAGTGT	426
QY	61	CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA	120
Db	427	CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA	486
QY	121	TGGAGATCAACACCGGGGACTTCTCCACACAGGCTCAACCATGTTACAACTCC	180
Db	487	TGGAGATCAACACCGGGGACTTCTCCACACAGGCTCAACCATGTTACAACTCC	546
QY	181	GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT	240
Db	547	GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTANANAAAGCCCTGGT	606
QY	241	GCAGCGGCTTGTCTGGGGATGTGAGCGCTCAGGAGTATGATGAGTACTCGTGGTCTTCGG	300

Db 607 GCANGCGGCTTCTGGGGATGTGAGCGCTCANGAGTGATNAGTACTCTGGTCTGG 666

RESULT 8

HS1100H13/c

LOCUS

DEFINITION

HS1100H13 116792 bp DNA linear PRI 18-MAR-2001  
Human DNA sequence from clone RP5-1100H13 on chromosome 20q11.2  
Contains the 3' end of gene KIAA1219, a putative novel gene, a DC5  
pseudogene, the gene for a putative RhoGAP domain containing  
protein, ESTs, STSS, GSSs and two CpG islands, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL035419.12 GI:11546045  
HTG: CpG Island; DC5; KIAA1219; RhoGAP.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

1 (bases 1 to 116792)  
Smith,M.

Direct Submission

Submitted (12-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerquest@sanger.ac.uk

COMMENT

On Dec 5, 2000 this sequence version replaced gi:10198625.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/Celegans/wormpep> This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr20>

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP5-1100H13 is from  
the library RPCI-5 constructed by the group of Pieter de Jong. For  
further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP5-1100H13 The true  
left end of clone RP11-12201 is at 47827 in this sequence. The true  
right end of clone RP5-927M24 is at 16404 in this sequence.

FEATURES

source

Location/Qualifiers  
1..116792

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="R2PD:RPCIP704H131100"

/db\_xref="taxon:9606"

/chromosome="20"

/map="q11.2"

/clone="RP5-1100H13"

/clone\_lib="RPCI-5"

repeat\_region 1331..1465

/note="MER58C repeat: matches 68..89 of consensus"

gene

1587..13349

/gene="dJ1100H13.1"

join(<1587..1723,4379..4487,5260..5338,8641..8789,

mRNA

9265..13349)

/gene="dJ1100H13.1"

/product="dJ1100H13.1.1 (KIAA1219 (similar to Drosophila  
G09358 and C. elegans D2085.5) isoform 1)"  
/note="match: cDNAs: Em:AB033045 Em:AL132998"

```

match: ESTs: Em:AI190579 Em:AA256927 Em:N36561 Em:AA874351
Em:AI281190 Em:AA180094 Em:AA669533 Em:AA511971
Em:AI193307 Em:AA215757 Em:AA156777 Em:AA131113
Em:AA418776 Em:AI357456 Em:AI379408 Em:N76170 Em:W52362
Em:AA024901 Em:AI334294 Em:AA205264 Em:AA215603
Em:AA418686 Em:AA024902 Em:AI379162 Em:AA424825
Em:AI202970 Em:AI248059 Em:AA648752 Em:AI359832
Em:AI366165 Em:AI224862 Em:AI359759 Em:AI492050 Em:R76319
Em:AI050034 Em:AI423927 Em:AA024896 Em:R78702 Em:AI066457
Em:R72124 Em:AI361138 Em:R81299 Em:AA282907 Em:L44403
Em:AI500502 Em:Z98499 Em:R83793 Em:AA412128 Em:AA179405
Em:AA459722 Em:AA173335 Em:AA283091 Em:H26506 Em:H02652
Em:AA258043 Em:N64761 Em:AA255436 Em:AI271673 Em:AX017681
Em:AA95863 Em:AI221344 Em:R65916 Em:W07865 Em:AA383163
Em:H26520 Em:AA872554 Em:AA343687 Em:AA889796 Em:R68642
Em:AA580081 Em:AA156671 Em:AA552080 Em:AA024895"
/evidence=not_experimental
join(<1587. .1723,4379. .4487,5260. .5338,8641. .8789,
9265. .9458)
/gene="dJ1100H13.1"
/note="Continues in Em:AL049868 as dJ927M24.2
match: proteins: Tr:Q9Y3G9 Tr:Q9W2x2 Tr:Q9U4G6 Tr:Q18431"
/codon_start=3
/evidence=not_experimental
/product="dJ1100H13.1.1 (K1AA1219 (similar to Drosophila
GH03358 and C. elegans D2085.5) isoform 1)"
/protein_id="CAB41646.1"
/db_xref="GI:4704208"
/db_xref="SPTREMBL:Q9Y3G9"
/translation="DSLESNISDSDSDSNMDLPGILKQPSLTLELPNHTDNLSSQ
RLSPSSMRKLQPGRPVPLGPETRVSVVWVYRDDIENPLSELMTFISGVETTAN
SSTLSRSTLEKEPVLFILHPLNTGLERIKQATGFENNVIPLDGMIVSRALGEL
VRQIVINCRRKLESISYPHVRKQKIIDIYNKRNKQLEPFYTSUFQVGLKN
CSS"
2753. .3187
/note="L1M4 repeat: matches 3867. .4299 of consensus"
4727. .5189
/gene="dJ1100H13.1"
/note="match: GSS: Em:A0663797"
4732. .4875
/gene="dJ1100H13.1"
/note="match: STS: Em:HSPF04E1"
4745. .5581
/gene="dJ1100H13.1"
/note="match: GSS: Em:AQ745415"
complement(4997. .5442)
/note="match: STS: Em:G33216"
complement(5210. .5903)
/note="match: GSS: Em:AQ321022"
5522. .6464
/note="L1MA10 repeat: matches 5388. .6322 of consensus"
6837. .6970
/note="MIR repeat: matches 32. .175 of consensus"
complement(6984. .7365)
/note="match: GSS: Em:AQ807920"
7232. .7391
/note="L1MC4 repeat: matches 7614. .7783 of consensus"
7610. .7684
/note="AluJb repeat: matches 39. .113 of consensus"
join(<8150. .8789,9265. .9431,10364. .>10484)
/gene="dJ1100H13.1"
/product="dJ1100H13.1.2 (K1AA1219, isoform 2)"
/note="match: ESTs: Em:AL536595"
/evidence=not_experimental
8567. .8616
/note="2 copies 25 mer 98% conserved"
12187. .12234
/note="12 copies 4 mer caca 89% conserved"
12536. .12571
/note="18 copies 2 mer aa 80% conserved"
12911. .12955
/note="L2 repeat: matches 2657. .2701 of consensus"
complement(12943. .13346)

```

CDS

```

polyA_signal
polyA_site
repeat_region
gene
mRNA
mRNA
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
polyA_site
mRNA
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

```

Query Match  
Best Local Similarity 95.7%; DB 9; Length 116792;  
Matches 287; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```

QY 1 ACGAAATCCGACCCCTGGTCAAGGATATATGGGACATCGTATAGCCAACTCGAGTGT 60
Db 49629 ACGAAATCCGACCCCTGGTCAAGGATATATGGGACATCGTATAGCCAACTCGAGTGT 49570
QY 61 CTGCTGACAGCTTTGTGAGACACGAGGACACATGCGAAGCTGGAACCTTGCACCTTGA 120
Db 49569 CTGCTGACAGCTTTGTGAGACACGAGGACACATGCGAAGCTGGAACCTTGCACCTTGA 49510
QY 121 TGGAGATCAACACCGAGGAGCTTTCTCTACACAAGCGCTCAACACATGTACAACTCC 180

```

||||| 49509 TGGAGATCAACCCAGCGGACTTCTCACACAAGCGCTCAACACACGTACAAACT-C 49451  
||||| 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGGT 240  
||||| 49450 GCACAAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGGT 49391  
||||| 241 GCAGGCGGCTTGGTGGGGATGTGACCGCTCAGGAGCTGATGAGGACTCAGGCTGGT 300  
||||| 49390 GCAGGCGGCTTGGTGGGGATGTGA-CGCTCAGGAGCTGATGAGGACTCAGGCTGGT 49332  
RESULT 9  
AC009068/c  
LOCUS AC009068 160356 bp DNA linear PRI 27-APR-2001  
DEFINITION Homo sapiens chromosome 16 clone RP11-314K3, complete sequence.  
AC009068  
VERSION AC009068.10 GI:13811892  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 160356)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL Direct Submission  
REFERENCE 2 (bases 1 to 160356)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Direct Submission  
REFERENCE 3 (bases 1 to 160356)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
REFERENCE 4 (bases 1 to 160356)  
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
COMMENT On Apr 27, 2001 this sequence version replaced gi:12000278.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
Quality: Phrap Quality >=40 99.6% of Sequence  
Estimated Total Number of Errors is 0.3.  
STS Content:  
WI-11796 G24048  
SHGC-37143 G30481.  
FEATURES  
source  
1. 160356  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-314K3"  
BASE COUNT 37469 a 43134 c 43308 g 36445 t  
ORIGIN  
Query Match 67.0%; Score 201; DB 9; Length 160356.  
Best Local Similarity 100.0%; Pred. No. 3.9e-43;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 100 AGCTGATACCTTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 159  
DB 50029 AGCTGATACCTTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 159  
QY 160 TCAACACATGTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 219  
DB 49969 TCAACACATGTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 49910  
QY 220 ACTTCTAGAGAAAGGCTTGGTGGGGATGTGAGCGCTCAGGAGCGTG 279  
DB 49909 ACTTCTAGAGAAAGGCTTGGTGGGGATGTGAGCGCTCAGGAGCGTG 49850

280 ATGAGGTACTCGTGGTTCTGG 300  
||||| 49849 ATGAGGTACTCGTGGTTCTGG 49829  
RESULT 10  
AC123908/c  
LOCUS AC123908 185321 bp DNA linear PRI 28-MAR-2003  
DEFINITION Homo sapiens chromosome 16 clone CTD-2542L18, complete sequence.  
AC123908  
VERSION AC123908.3 GI:29336199  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 185321)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL Direct Submission  
REFERENCE 2 (bases 1 to 185321)  
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los  
TITLE Direct Submission  
JOURNAL Direct Submission  
REFERENCE 3 (bases 1 to 185321)  
AUTHORS Alamos National Laboratory.  
TITLE Direct Submission  
JOURNAL Submitted (06-SEP-2002) Production Sequencing Facility, DOE Joint  
REFERENCE 4 (bases 1 to 185321)  
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center and Los  
JOURNAL Direct Submission  
COMMENT On Mar 28, 2003 this sequence version replaced gi:22748376.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center and Los Alamos  
National Laboratory  
Quality: Phrap Quality >=40 99.9% of Sequence  
Estimated Total Number of Errors is 0.1.  
FEATURES  
source  
1. 185321  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2542L18"  
BASE COUNT 40674 a 50904 c 51869 g 41874 t  
ORIGIN  
Query Match 67.0%; Score 201; DB 9; Length 185321;  
Best Local Similarity 100.0%; Pred. No. 3.8e-43;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 100 AGCTGATACCTTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 159  
DB 95218 AGCTGATACCTTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 159  
QY 160 TCAACACATGTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 219  
DB 95158 TCAACACATGTACAACTCCGACCACTCCAGCTCTGGAGAGTACTCAGTCTCAGG 95099  
QY 220 ACTTCTAGAGAAAGGCTTGGTGGGGATGTGAGCGCTCAGGAGCGTG 279  
DB 95098 ACTTCTAGAGAAAGGCTTGGTGGGGATGTGAGCGCTCAGGAGCGTG 95039  
QY 280 ATGAGGTACTCGTGGTTCTGG 300  
|||||



Db 95038 ATGAGTACTCGGTCTGG 95018

RESULT 11

AC137771

LOCUS

DEFINITION

AC137771 Homo sapiens chromosome 16 clone CTD-2139A24, WORKING DRAFT

SEQUENCE, 45 unordered pieces.

ACCESSION

AC137771

VERSION

AC137771.1 GI:26006513

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 245210)

DOE Joint Genome Institute.

TITLE

Sequencing of Human Chromosome 16

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 245210)

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (03-DEC-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

AUTHORS

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

-----

Project Information

Center Project Name: 680949

Center clone name: CITB-HL\_2139A24

-----

Summary Statistics

Consensus quality: 221831 bases at least Q40

Consensus quality: 226530 bases at least Q30

Consensus quality: 229658 bases at least Q20

Estimated insert size: 170000; agarose-fp estimation

Estimated insert size: 240810; sum-of-contigs estimation

Quality coverage: 17.31 in Q20 bases; agarose-fp estimation

Quality coverage: 12.22 in Q20 bases; sum-of-contigs estimation

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 45 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1 1084: contig of 1084 bp in length

1085 1184: gap of unknown length

1185 2248: contig of 1064 bp in length

2249 2348: gap of unknown length

2349 3435: contig of 1087 bp in length

3436 3535: gap of unknown length

3536 4637: contig of 1102 bp in length

4638 4737: gap of unknown length

4738 5816: contig of 1079 bp in length

5817 5916: gap of unknown length

5917 6951: contig of 1035 bp in length

6952 7051: gap of unknown length

7052 8599: contig of 1548 bp in length

8600 8699: gap of unknown length

8700 10508: contig of 1809 bp in length

10509 10609: gap of unknown length

10609 11838: contig of 1130 bp in length

11739 11839: gap of unknown length

13754: contig of 1916 bp in length

13755 13854: gap of unknown length

13855 14996: contig of 1142 bp in length

14997 15096: gap of unknown length

15097 16852: contig of 1756 bp in length

16853 16952: gap of unknown length

16953 18502: contig of 1550 bp in length

18503 18602: gap of unknown length

18603 20736: contig of 2134 bp in length

20737 20836: gap of unknown length

20837 22243: contig of 1407 bp in length

22244 22343: gap of unknown length

22344 24310: contig of 1967 bp in length

24311 24411: gap of unknown length

24411 26675: contig of 2265 bp in length

26676 26775: gap of unknown length

26776 28257: contig of 1482 bp in length

28258 28357: gap of unknown length

28358 29668: contig of 1311 bp in length

29669 31503: contig of 1735 bp in length

31504 31603: gap of unknown length

31604 33962: contig of 2359 bp in length

33963 34062: gap of unknown length

34063 36099: contig of 2037 bp in length

36100 38832: contig of 2633 bp in length

38833 38932: gap of unknown length

38933 41464: contig of 2532 bp in length

41465 41564: gap of unknown length

41565 44085: contig of 2521 bp in length

44086 44185: gap of unknown length

44186 47029: contig of 2843 bp in length

47029 47129: gap of unknown length

47129 49395: contig of 2267 bp in length

49396 49495: gap of unknown length

49496 51435: contig of 1940 bp in length

51436 51535: gap of unknown length

51536 55707: contig of 4171 bp in length

55707 55806: gap of unknown length

55807 60092: contig of 4286 bp in length

60093 60192: gap of unknown length

60193 63448: contig of 3156 bp in length

63449 66966: contig of 3518 bp in length

66967 67066: gap of unknown length

67067 71746: contig of 4680 bp in length

71747 71846: gap of unknown length

71847 74430: contig of 2584 bp in length

74431 74530: gap of unknown length

74531 79515: contig of 4985 bp in length

79516 83377: gap of unknown length

83378 83477: gap of unknown length

83478 88208: contig of 4731 bp in length

88209 88308: gap of unknown length

88309 93320: contig of 5012 bp in length

93321 93420: gap of unknown length

93421 97345: contig of 3925 bp in length

97346 97445: gap of unknown length

97446 100893: contig of 3448 bp in length

100894 100993: gap of unknown length

100994 107097: contig of 6104 bp in length

107098 107197: gap of unknown length

107198 111963: contig of 4766 bp in length

111964 112063: gap of unknown length

112064 116267: contig of 4204 bp in length

116268 116367: gap of unknown length

116368 123219: contig of 6852 bp in length

123220 123320: gap of unknown length

123320 245210: contig of 121891 bp in length.

Location/Qualifiers

1..245210

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="16"

/clone="CTD-2139A24"

/clone\_lib="Caltech human BAC library D"

FEATURES

Source



BASE COUNT 62851 a 55678 c 56421 g 65631 t 4629 others  
ORIGIN

Query Match 67.0%; Score 201; DB 2; Length 245210;

Best Local Similarity 100.0%; Pred. No. 3.6e-43;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 ACCTGGATTAACCTTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCACAAGCGC 159

Db 115926 ACCTGGATTAACCTTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCACAAGCGC 115985

QY 160 TCAACACATGTACAACTCCGACGACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGG 219

Db 115986 TCAACACATGTACAACTCCGACGACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGG 116045

QY 220 ACTTCTAGAGAAGCGCTGGTGCAGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTG 279

Db 116046 ACTTCTAGAGAAGCGCTGGTGCAGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTG 116105

QY 280 ATGAGGTACTCGTGGTCTGG 300

Db 116106 ATGAGGTACTCGTGGTCTGG 116126

# RESULT 12

AK091519

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

AK091519 Homo sapiens cDNA FLJ34200 fis, clone FCBF3019663. PRI 15-JUL-2002

AK091519 Homo sapiens (human) (full insert sequence).

AK091519.1 GI:21749909

Oligo capping; fis (full insert sequence).

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,

Ono, Y., Hotata, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,

Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,

Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,

Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,

Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K.,

Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M.,

Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,

Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 3097)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and

Evaluation; clone selection for full insert sequencing: HRI and

RAB; annotation: HRI and RAB.

Location/Qualifiers

1..3097

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="FCBF3019663"

/tissue\_type="brain"

/clone\_lib="FCBF3"

/dev\_stage="fetal"

/note="cloning vector: pME18SFL3"

726 a 703 c 786 g 882 t

BASE COUNT 62851 a 55678 c 56421 g 65631 t 4629 others

ORIGIN

Query Match 66.5%; Score 199.4; DB 9; Length 3097;

Best Local Similarity 99.5%; Pred. No. 2.1e-42;

Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 100 ACCTGGATTAACCTTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCACAAGCGC 159

Db 2734 ACCTGGATTAACCTTGACCTTGATGGAGATCAACACAGCGGGACTTTCCTCACAAGCGC 2793

QY 160 TCAACACATGTACAACTCCGACGACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGG 219

Db 2794 TCAACACATGTACAACTCCGACGACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGG 2853

QY 220 ACTTCTAGAGAAGCGCTGGTGCAGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTG 279

Db 2854 ACTTCTAGAGAAGCGCTGGTGCAGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTG 2913

QY 280 ATGAGGTACTCGTGGTCTGG 300

Db 2914 ATGAGGTACTCGTGGTCTGG 2934

RESULT 13

AX336730

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match 46.7%; Score 140; DB 6; Length 454;

Best Local Similarity 99.3%; Pred. No. 1.8e-26;

Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60

Db 303 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362

QY 61 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 120

Db 363 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 421

QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACA 152

Db 422 TGGAGATCAACACAGCGGGACTTTCCTCACA 453

RESULT 14

AB097169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Query Match 93.3%; Score 933 bp DNA

Best Local Similarity 99.3%; Pred. No. 1.8e-26;

Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60

Db 303 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362

QY 61 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 120

Db 363 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 421

QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACA 152

Db 422 TGGAGATCAACACAGCGGGACTTTCCTCACA 453

RESULT 14

AB097169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Query Match 93.3%; Score 933 bp DNA

Best Local Similarity 99.3%; Pred. No. 1.8e-26;

Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60

Db 303 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362

QY 61 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 120

Db 363 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 421

QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACA 152

Db 422 TGGAGATCAACACAGCGGGACTTTCCTCACA 453

RESULT 14

AB097169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Query Match 93.3%; Score 933 bp DNA

Best Local Similarity 99.3%; Pred. No. 1.8e-26;

Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60

Db 303 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362

QY 61 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 120

Db 363 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 421

QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACA 152

Db 422 TGGAGATCAACACAGCGGGACTTTCCTCACA 453

RESULT 14

AB097169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Query Match 93.3%; Score 933 bp DNA

Best Local Similarity 99.3%; Pred. No. 1.8e-26;

Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60

Db 303 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362

QY 61 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 120

Db 363 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 421

QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACA 152

Db 422 TGGAGATCAACACAGCGGGACTTTCCTCACA 453

RESULT 14

AB097169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Query Match 93.3%; Score 933 bp DNA

Best Local Similarity 99.3%; Pred. No. 1.8e-26;

Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60

Db 303 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362

QY 61 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 120

Db 363 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 421

QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACA 152

Db 422 TGGAGATCAACACAGCGGGACTTTCCTCACA 453

RESULT 14

AB097169

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Query Match 93.3%; Score 933 bp DNA

Best Local Similarity 99.3%; Pred. No. 1.8e-26;

Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60

Db 303 ACGAATCCGGACCTGGTCAAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 362

QY 61 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 120

Db 363 CTGCTCAGAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGGATTAACCTTGCACCTTGA 421

QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACA 152

Db 422 TGGAGATCAACACAGCGGGACTTTCCTCACA 453

RESULT 14

AB097169

LOCUS

DEFINITION

ACCESSION

VERSION

ORGANISM	Xenopus laevis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.
AUTHORS	1 Kubota, Y., Takase, Y., Komori, Y., Hashimoto, Y., Arata, T., Kanimura, Y., Araki, H. and Takisawa, H.
TITLE	A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication
JOURNAL	Genes Dev. 17 (9), 1141-1152 (2003)
MEDLINE	22615695
PUBMED	12730133
REFERENCE	2 (bases 1 to 933)
AUTHORS	Kubota, Y. and Takisawa, H.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2002) Yumiko Kubota, Osaka University, Department of Biology, Graduate School of Science, Machikaneyamacho 1-1, Toyonaka, Osaka 560-0043, Japan
FEATURES	1 (E-mail: ykubota@bio.sci.osaka-u.ac.jp, Tel: 81-6-6850-5554, Fax: 81-6-6850-5554)
source	Location/Qualifiers
gene	1..933
CDS	/organism="Xenopus laevis" /mol_type="genomic DNA" /db_xref="taxon:8355" 21..578 /gene="Psf2" 21..578 /gene="Psf2" /function="essential for DNA replication" /codon_start=1 /product="Psf2" /protein_id="BAC66459.1" /db_xref="GI:29365481" /translation="MDASVEFLAEKQVTVIPNSLDKVLIGDLPENPSLPVEV PLWLAINLKKORCKIVPEWMDVEKLIARDQERETFTFMPSPYMETKLILNH AADNTPKADEIETLVKDTWDTRIAKIRLSADSFVKGQEAHKLNLTLMEINTIGTFE TSLHNKYLKRLSILNPEEGSQDIY"
BASE COUNT	282 a 200 c 208 g 243 t
ORIGIN	
Query Match	41.9%; Score 125.8; DB 5; Length 933;
Best Local Similarity	68.1%; Pred. No. 9.2e-23;
Matches	175; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY	1 AGCAATCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60
Db	352 ATGAGATTGCTACCTTGTATTAAGACACTGGGATACAAAGTAGCAAACTGGCTCT 411
QY	61 CTGCTGACAGCTTGTGACAGCAGGAGGACACTGCCAAGCTGGATTAACCTTGACCTTGA 120
Db	412 CTGCTGACAGCTTGTGAAAGGGCAGGAGCTCATGCCAAGCTGGATTAACCTGACCTAA 471
QY	121 TGGAGATCAACACCGCGGACTTTTCCTCACACAGCGCTCAACACATGTACAAACTCC 180
Db	472 TGGAAATTAACACCTCGGACATTTTACTAGTCTTTTAAACCATGTACAGCTAC 531
QY	181 GCACGAACTCCAGCCTCTGGAGAGTACTACTCTCAGGACTCTTAGAAGAGCCCTGGT 240
Db	532 GTACCAAGCTGCAACCAAGGAGGACACTCAGAGGATTAACGCAATTGCTGCA 591
QY	241 GCAGGCGGCTGCTGGG 257
Db	592 ATTATCTGGCTGCTGAG 608
RESULT	15
AC117085/c	
LOCUS	AC117085 207307 bp DNA linear HTG 13-MAY-2003
DEFINITION	Rattus norvegicus clone CH230-247G13, WORKING DRAFT SEQUENCE.
ACCESSION	AC117085
VERSION	AC117085.8 GI:30579062
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GMGE  
Center clone name: CH230-247G13  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 198554 bases at least Q40  
Consensus quality: 200661 bases at least Q30  
Consensus quality: 201900 bases at least Q20  
Estimated insert size: 210626; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length .  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\*  
\* 1 207307: contig of 207307 bp in length.

FEATURES  
source  
1..207307  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-247G13"  
87039..177061  
/note="clone\_boundary"  
clone\_end:Sp6  
site:MboI  
end\_sequence:BZ168551"  
150202..150357  
/note="clone\_boundary"  
clone\_end:T7  
site:MboI  
end\_sequence:BZ168550"  
BASE COUNT 43941 a 54080 c 55681 g 48886 t 4719 others  
ORIGIN

Query Match 37.3%; Score 111.8; DB 2; Length 207307;  
Best Local Similarity 78.4%; Pred. NO. 1.9e-19;  
Matches 134; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
QY 100 AGCTGGATAACCTTACCTTGATGGAGATCAACACAGCGGACCTTCCTCACAAAGCGC 159  
Db 196735 AGCTGGACACCTTACCTTGATGGAGATCAACACAGCGGACCTTCCTCACAAAGCGC 196676  
QY 160 TCAACACATCTACAACTCCGACGACCTCCAGCCCTCTCGAGAGTACTCAGTCTCAGG 219  
Db 196675 TCAACACATCTACAACTCCGACGACCTTCAGCCCTCTCGAGAGTACTCAGTCTCAGG 196616  
QY 220 ACTTCTAGAGAAAGCCCTGGTGCGAGGGGCTTGCTGGGGGATGTGAGCGCT 270  
Db 196615 ACTTCTAGCCAGAGCCCTCTGGTCTGCNAACCCCTTGCTGGGTTAGAGTGAT 196565

Search completed: October 14, 2003, 17:02:25  
Job time : 1485 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 11:32:16 ; Search time 1272.67 Seconds  
(without alignments)  
5729.185 Million cell updates/sec

Title: US-09-854-124-5  
Perfect score: 300  
Sequence: 1 acgaataccgaccctgtc.....tgaggtaactcgtgtctcgg 300

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estinu:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: em\_gss\_hum:\*
  - 18: em\_gss\_inv:\*
  - 19: em\_gss\_pln:\*
  - 20: em\_gss\_fun:\*
  - 21: em\_gss\_fun:\*
  - 22: em\_gss\_mam:\*
  - 23: em\_gss\_mus:\*
  - 24: em\_gss\_pro:\*
  - 25: em\_gss\_rod:\*
  - 26: em\_gss\_phg:\*
  - 27: em\_gss\_vrl:\*
  - 28: gb\_gsl1:\*
  - 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	300	100.0	533	14	CB112523 K-EST0154
C 2	300	100.0	559	12	BM126492 lf06h06.y
C 3	300	100.0	659	13	BU729618 01-E-COI-
C 4	300	100.0	716	10	BE382866 601297755

5	300	100.0	801	12	BI196248
6	300	100.0	876	10	BG181130
7	300	100.0	896	10	BE562088
8	300	100.0	914	10	BE547290
9	300	100.0	943	10	BF683514
10	300	100.0	970	9	AL560669
11	300	100.0	1007	12	BM449472
12	300	100.0	1064	13	BQ277667
13	300	100.0	1071	13	BU184963
14	300	100.0	1201	9	AL560880
15	300	100.0	1201	9	AL582250
16	299.6	99.9	1201	9	AL582217
17	299	99.7	810	12	BQ009838
18	298.8	99.6	1193	9	AL529785
19	298.4	99.5	488	14	CB160336
20	298.4	99.5	696	9	AW249012
21	298.4	99.5	712	13	BQ625683
22	298.4	99.5	714	10	BG104289
23	298.4	99.5	716	12	BG825252
24	298.4	99.5	730	10	BE296429
25	298.4	99.5	790	10	BE795306
26	298.4	99.5	866	10	BF795157
27	298.4	99.5	867	13	BQ233393
28	298.4	99.5	891	11	BQ228339
29	298.4	99.5	921	10	BE799212
30	298.4	99.5	932	10	BE796384
31	298.4	99.5	1028	10	BE561044
32	298.4	99.5	1030	10	BF311745
33	298.4	99.5	1182	11	AF125098
34	298	99.3	833	13	BU597296
35	297.4	99.1	494	14	H64860
36	296.8	98.9	895	13	BU539659
37	293.2	97.7	1056	13	BA456910
38	292.4	97.5	667	9	AU126087
39	292	97.3	966	9	AL582077
40	291.4	97.1	1201	9	AL562756
41	290	96.7	849	10	BE514071
42	289.6	96.5	1007	10	BE795838
43	288.6	96.2	1201	9	AL526847
44	288	96.0	761	12	BI257993
45	287.4	95.8	966	10	BE791539

ALIGNMENTS

RESULT 1  
CB112523/c  
LOCUS K-EST0154352 L6ChoCK0 Homo sapiens cDNA clone L6ChoCK0-4-D04 5',  
DEFINITION mRNA sequence.  
ACCESSION CB112523  
VERSION CB112523.1 GI:27938330  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 533)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 4 row: D column: 04  
High quality sequence stop: 533.

FEATURES	Location/Qualifiers	TITLE
source	1..533	JOURNAL
	/organism="Homo sapiens"	COMMENT
	/mol_type="mRNA"	

```

Location/Qualifiers
1. 533
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L6ChoCK0-4-D04"
/sex="M"
/cell_line="Cho-CK"
/lab_host="Top10F"
/clone_lib="L6ChoCK0"
/Note="Organ: Liver; Vector: pCNS-D2; Site.1: EcoRI;
Site.2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
109 a 143 c 135 g 146 t

```

[illegible]

TITLE : Jackson, Y. and Bowers, Y.  
 JOURNAL : Endocrine Pancreas Consortium  
 COMMENT : Unpublished  
 Other\_ESTS : if06h06.x1  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@molb.harvard.edu  
 Library was constructed by Dr. Douglas Melton DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Juliana Brown  
 (brownetas.harvard.edu) This sequence now available from the IMAGE  
 consortium, for clone orders contact: info@image.llnl.gov  
 High quality sequence stop: 437.

FEATURES  
SOURCE

```

BASE COUNT      157 a      131 c      142 g      129 t
ORIGIN
Query Match      100.0%; Score 300; DB 12; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.8e-72;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1  ACAGAAATCCGACCCTTGTC AAGGATATGTGGGACACTCGTATAGCCAAAACCTCCGAGTGT 60
|||||
118 ACAGAAATCCGACCCTTGTC AAGGATATGTGGGACACTCGTATAGCCAAAACCTCCGAGTGT 177
|||||

61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 120
|||||

178 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGACCTTGA 237
|||||

121 TGGAGATCAACACAGCGGGGACTTTCCCTCACAAGGGCTCAACACATGTACAAACTCC 180
|||||

238 TGGAGATCAACACAGCGGGGACTTTCCCTCACAAGGGCTCAACACATGTACAAACTCC 297
|||||

181 GCAGGAACCTCCAGCCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAGGCCCTGGT 240
|||||

298 GCAGGAACCTCCAGCCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAGGCCCTGGT 357
|||||

241 GCAGGCGGCTTGCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCCTGTGTTCTGG 300
|||||

358 GCAGGCGGCTTGCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCCTGTGTTCTGG 417
|||||

RESULT 3
107729618/c

```

LOCUS BU729618 659 bp mRNA linear EST 09-OCT-2002  
 DEFINITION UI-E-CQ1-asy-b-18-0-UI.s1 UI-E-CQ1 Homo sapiens cDNA clone  
 UI-E-CQ1-asy-b-18-0-UI 3', mRNA sequence.  
 ACCESSION BU729618  
 VERSION BU729618.1 GI:23652681  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 659)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 FORWARD  
 POLYA-Yes.

## FEATURES

Location/Qualifiers  
 1..659  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-CQ1-asy-b-18-0-UI"  
 /tissue\_type="optic nerve"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-CQ1"  
 /note="Organ: eye; Vector: p7T3-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
 UI-E-CQ1 is a normalized cDNA library containing the  
 following tissue(s): optic nerve. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into p7T3-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (AT)18 tail. The sequence tag for this library is  
 CCAATTAAGG. This library was created for the program, Gene  
 Discovery in the Visual System, supported by National Eye  
 Institute (NEI).  
 TAG\_LIB=UI-E-CQ1  
 TAG\_TISSUE=human optic nerve  
 TAG\_SEQ=CCATTAAAGT"

BASE COUNT 149 a 159 c 151 g 200 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 13; Length 659;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 Db 509 ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 450  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAANGCTGGATTAACCTTGACCTTGA 120

Db 449 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGACCTTGA 390  
 QY 121 TGGAGATCAACACACAGCGGACTTCTCCTCACACAAGCGCTCAACACACATGTACAAACTCC 180  
 Db 389 TGGAGATCAACACACAGCGGACTTCTCCTCACACAAGCGCTCAACACACATGTACAAACTCC 330  
 QY 181 GCACGAACTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
 Db 329 GCACGAACTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 270  
 QY 241 GCAGGCGGCTTCTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 300  
 Db 269 GCAGGCGGCTTCTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 210

## RESULT 4

BE382866 716 bp mRNA linear EST 21-JUL-2000  
 LOCUS 601297755F1 NTH\_MGC\_19 Homo sapiens cDNA clone IMAGE:3627852 5',  
 DEFINITION mRNA sequence.  
 BE382866  
 ACCESSION BE382866.1 GI:9328231  
 VERSION BE382866  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 716)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM311 row: n column: 13  
 High quality sequence stop: 651.

## FEATURES

Location/Qualifiers  
 1..716  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3627852"  
 /tissue\_type="neuroblastoma"  
 /clone\_lib="NIH\_MGC\_19"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

BASE COUNT 185 a 179 c 195 g 157 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 10; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 2e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 Db 337 ACGAAATCCGGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 396  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGACCTTGA 120  
 Db 397 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATTAACCTTGACCTTGA 456

QY 121 TGGAGATCAACACCGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 180  
 |||||  
 Db 457 TGGAGATCAACACCGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 516  
 |||||  
 QY 181 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCGTGGT 240  
 |||||  
 Db 517 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCGTGGT 576  
 |||||  
 QY 241 GCAGCGCGCTTCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGCTGTTCTGG 300  
 |||||  
 Db 577 GCAGCGCGCTTCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGCTGTTCTGG 636  
 |||||

RESULT 5  
 B1196248  
 LOCUS 801 bp mRNA linear EST 10-JUL-2001  
 DEFINITION 602754709F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4890147 5',  
 mRNA sequence.  
 ACCESSION B1196248  
 VERSION B1196248.1 GI:14651268  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 TITLE NIH-MGC http://mgc.nci.nih.gov/.  
 AUTHORS 1 (bases 1 to 801)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LICM1767 row: b column: 04  
 High quality sequence stop: 781.

## FEATURES

Location/Qualifiers  
 1..801  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4890147"  
 /tissue\_type="neuroblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_19"  
 /note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; CDNA made by oligo-dr priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCAGAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 208 a 196 c 216 g 181 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 12; Length 801;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAAATCCGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 |||||  
 Db 356 ACAGAAATCCGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 415  
 |||||  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAAGCTGGATAAAGTTGACCTTGA 120  
 |||||  
 Db 416 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAAGCTGGATAAAGTTGACCTTGA 475  
 |||||  
 QY 121 TGGAGATCAACACCGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 180

Db 476 TGGAGATCAACACCGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 535  
 |||||  
 QY 181 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCGTGGT 240  
 |||||  
 Db 536 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCGTGGT 595  
 |||||  
 QY 241 GCAGCGCGCTTCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGCTGTTCTGG 300  
 |||||  
 Db 596 GCAGCGCGCTTCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGCTGTTCTGG 655  
 |||||

## RESULT 6

BG181130  
 LOCUS 876 bp mRNA linear EST 06-FEB-2001  
 DEFINITION 602329264F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4430823 5',  
 mRNA sequence.  
 ACCESSION BG181130  
 VERSION BG181130.1 GI:12687833  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 TITLE NIH-MGC http://mgc.nci.nih.gov/.  
 AUTHORS 1 (bases 1 to 876)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTP  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10184 Row: 0 Column: 16  
 High quality sequence stop: 674.

## FEATURES

Location/Qualifiers  
 1..876  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4430823"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_91"  
 /note="Organ: prostate; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dr primed.  
 Average insert size 1.4 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 246 a 217 c 214 g 199 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 10; Length 876;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAAATCCGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 |||||  
 Db 115 ACAGAAATCCGACCCCTGGTCAAGGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 174  
 |||||  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAAGCTGGATAAAGTTGACCTTGA 120  
 |||||  
 Db 175 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAAGCTGGATAAAGTTGACCTTGA 234  
 |||||  
 QY 121 TGGAGATCAACACCGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 180  
 |||||  
 Db 235 TGGAGATCAACACCGCGGGACTTTCCTCACACAAGCGCTCAACACATGTACAAACTCC 294  
 |||||  
 QY 181 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCGTGGT 240



Db 295 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 354  
 QY 241 GCAGGCGGCTTCTGGGGGATGTGAGCGCTCAGGACGCTGATGAGGTACTCGTGGTCTCGG 300  
 Db 355 GCAGGCGGCTTCTGGGGGATGTGAGCGCTCAGGACGCTGATGAGGTACTCGTGGTCTCGG 414

RESULT 7  
 BE562088  
 LOCUS 601345038F1 NIH\_MGC\_8 896 bp mRNA linear EST 15-AUG-2000  
 DEFINITION mRNA sequence.  
 ACCESSION BE562088  
 VERSION BE562088.1 GI:9805808  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 896)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LCN354 row: k column: 09  
 High quality sequence start: 24  
 High quality sequence stop: 840.

FEATURES  
 source  
 1..896  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3678080"  
 /tissue\_type="Burkitt lymphoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_8"  
 /note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 218 a 230 c 251 g 197 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 10; Length 896;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAAATCCGACCTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 120  
 Db 438 ACGAAATCCGACCTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 497  
 QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 120  
 Db 498 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 557  
 QY 121 TGGAGATCAACACCGCGGACTTTCCCTACACAAGCGCTCAACACATGTACAACTCC 180  
 Db 558 TGGAGATCAACACCGCGGACTTTCCCTACACAAGCGCTCAACACATGTACAACTCC 617  
 QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 240

Query Match 100.0%; Score 300; DB 10; Length 896;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAAATCCGACCTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 60  
 Db 438 ACGAAATCCGACCTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 497  
 QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 120  
 Db 498 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 557  
 QY 121 TGGAGATCAACACCGCGGACTTTCCCTACACAAGCGCTCAACACATGTACAACTCC 180  
 Db 558 TGGAGATCAACACCGCGGACTTTCCCTACACAAGCGCTCAACACATGTACAACTCC 617  
 QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 240

Db 618 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 677  
 QY 241 GCAGGCGGCTTCTGGGGGATGTGAGCGCTCAGGACGCTGATGAGGTACTCGTGGTCTCGG 300  
 Db 678 GCAGGCGGCTTCTGGGGGATGTGAGCGCTCAGGACGCTGATGAGGTACTCGTGGTCTCGG 737

RESULT 8  
 BE547290  
 LOCUS 601073715F1 NIH\_MGC\_12 914 bp mRNA linear EST 09-AUG-2000  
 DEFINITION mRNA sequence.  
 ACCESSION BE547290  
 VERSION BE547290.1 GI:9775935  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 914)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM8453 row: m column: 03  
 High quality sequence stop: 658.

FEATURES  
 source  
 1..914  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3460010"  
 /tissue\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_12"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site:1: NotI;  
 Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."  
 BASE COUNT 254 a 227 c 238 g 195 t  
 ORIGIN

Query Match 100.0%; Score 300; DB 10; Length 914;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACGAAATCCGACCTTGTGACAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 60  
 Db 104 ACGAAATCCGACCTTGTGACAGGATATGTGGGACACTCGTATAGCCAACTCCGAGTGT 163  
 QY 61 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 120  
 Db 164 CTGCTGACAGCTTTGTGACAGCAGGAGGACATGCCAAGCTGGATACCTTGACCTTGA 223  
 QY 121 TGGAGATCAACACCGCGGACTTTCCCTACACAAGCGCTCAACACATGTACAACTCC 180  
 Db 224 TGGAGATCAACACCGCGGACTTTCCCTACACAAGCGCTCAACACATGTACAACTCC 283  
 QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 240  
 Db 284 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGCTGGT 343  
 QY 241 GCAGGCGGCTTCTGGGGGATGTGAGCGCTCAGGAGTACTCGTGGTCTCGG 300  
 Db 344 GCAGGCGGCTTCTGGGGGATGTGAGCGCTCAGGAGTACTCGTGGTCTCGG 403

Cervical  
 cell line

RESULT 9  
 LOCUS BF683514 943 bp mRNA linear EST 22-DEC-2000  
 DEFINITION 602139714F1 NIH\_MGC\_46 Homo sapiens cDNA clone IMAGE:4300947 5',  
 mRNA sequence.  
 ACCESSION BF683514  
 VERSION BF683514.1 GI:11968922  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 943)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LUCM1160 row: 1 column: 04  
 High quality sequence stop: 709.

FEATURES  
 source  
 1. 943  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4300947"  
 /tissue.type="leiomyosarcoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH-MGC\_46"  
 /note="Organ: uterus; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 214 a 253 c 272 g 204 t  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 10; Length 943;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACAGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 DB 369 ACAGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 428  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGAACCTTGA 120  
 DB 429 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGAACCTTGA 488  
 QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACAAAGCGCTCAACACATGTACAAATCC 180  
 DB 489 TGGAGATCAACACAGCGGGACTTTCCTCACAAAGCGCTCAACACATGTACAAATCC 548  
 QY 181 GCAGAACCTCCAGCCCTCGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
 DB 549 GCAGAACCTCCAGCCCTCGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 608  
 QY 241 GCAGGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGTTCTGG 300  
 DB 609 GCAGGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGTTCTGG 668

RESULT 10  
 LOCUS AL560669 970 bp mRNA linear EST 31-MAY-2003  
 DEFINITION AL560669 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CS0DL003YI21 5-PRIME, mRNA sequence.  
 ACCESSION AL560669  
 VERSION AL560669.2 GI:31284799  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 970)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12907354.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 7987.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DL003AE11Q1&cluster=7987.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DL003AE11Q1.

FEATURES  
 source  
 1. 970  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL003YI21"  
 /cell\_line="RAMOS CELL LINE"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 257 a 243 c 251 g 219 t  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 9; Length 970;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-72;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 DB 419 ACAGAAATCCGACCCCTGGTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 478  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGAACCTTGA 120  
 DB 479 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTGAACCTTGA 538  
 QY 121 TGGAGATCAACACAGCGGGACTTTCCTCACAAAGCGCTCAACACATGTACAAATCC 180  
 DB 539 TGGAGATCAACACAGCGGGACTTTCCTCACAAAGCGCTCAACACATGTACAAATCC 598  
 QY 181 GCAGAACCTCCAGCCCTCGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240  
 DB 599 GCAGAACCTCCAGCCCTCGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 658  
 QY 241 GCAGGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGTTCTGG 300  
 DB 659 GCAGGCGGCTTGTCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCTGTTCTGG 718

```

RESULT 11
BM449472
LOCUS
DEFINITION
  BM449472 1007 bp mRNA linear EST 05-FEB-2002
  5', mRNA sequence.
ACCESSION
  BM449472
KEYWORDS
  BM449472.1 GI:18498512
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 1007)
REFERENCE
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Life Technologies, Inc.
  DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: L1AM12117 row: m column: 08
  High quality sequence stop: 711.
FEATURES
  source
    1..1007
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5493679"
    /tissue_type="retinoblastoma"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH_MGC_67"
    /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
    Average insert size 1.75 kb. Library constructed by Life
    Technologies."
BASE COUNT
  259 a 254 c 262 g 231 t 1 others
  ORIGIN
    Query Match 100.0%; Score 300; DB 12; Length 1007;
    Best Local Similarity 100.0%; Pred. No. 2.4e-72;
    Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1 ACGAAATCCGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60
  DB 357 ACGAAATCCGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 416
  QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTTGACCTTGA 120
  DB 417 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTTGACCTTGA 476
  QY 121 TGGAGATCAACACAGCGGGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 180
  DB 477 TGGAGATCAACACAGCGGGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 536
  QY 181 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240
  DB 537 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 596
  QY 241 GCAGCGCGCTTGCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTCGG 300
  DB 597 GCAGCGCGCTTGCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTCGG 656

RESULT 12
BQ277667
LOCUS
DEFINITION
  BQ277667 1064 bp mRNA linear EST 07-MAY-2002
  5', mRNA sequence.
ACCESSION
  BQ277667
KEYWORDS
  BQ277667.1 GI:20487875
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 1064)
REFERENCE
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Ruben Laboratory
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLCM2042 row: d column: 19
  High quality sequence stop: 675.
FEATURES
  source
    1..1064
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5804514"
    /tissue_type="teratocarcinoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH_MGC_109"
    /note="Organ: ovary; Vector: pOTB7; Site_1: EcoRI; Site_2:
    XhoI; cDNA made by oligo-dT priming. Directionally cloned
    into EcoRI/XhoI sites using the following 5' adaptor:
    GGCACGAG(G). Library constructed by Ling Hong in the
    laboratory of Gerald M. Rubin (University of California,
    Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
    Superscript II RT (Life Technologies). Note: this is a
    NIH_MGC Library."
BASE COUNT
  275 a 268 c 273 g 245 t 3 others
  ORIGIN
    Query Match 100.0%; Score 300; DB 13; Length 1064;
    Best Local Similarity 100.0%; Pred. No. 2.5e-72;
    Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1 ACGAAATCCGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60
  DB 364 ACGAAATCCGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 423
  QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTTGACCTTGA 120
  DB 424 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCCAAGCTGGATAACTTTGACCTTGA 483
  QY 121 TGGAGATCAACACAGCGGGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 180
  DB 484 TGGAGATCAACACAGCGGGGACTTTCCTCACACAAGCGCTCAACACATGTACAACTCC 543
  QY 181 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 240
  DB 544 GCAGCAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGCCCTGGT 603
  QY 241 GCAGCGCGCTTGCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTCGG 300
  DB 604 GCAGCGCGCTTGCCTGGGGGATGTGAGCGCTCAGGACGTGATGAGGTACTCGTGGTCTCGG 663

RESULT 13
BU184963
LOCUS
DEFINITION
  BU184963 1071 bp mRNA linear EST 04-SEP-2002
  5', mRNA sequence.
ACCESSION
  BU184963

```

VERSION BUI84963.1 GI:22698947  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHC2269 row: m column: 13  
High quality sequence stop: 718.  
Location/Qualifiers  
1. .1071  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6002100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="PH109 (phage-resistant)"  
/clone\_lib="NIH\_MGC\_100"  
Note="Organ: liver; Vector: pOTB7; Site:1; XhoI; Site:2;  
ECORI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 283 a 262 c 275 g 246 t 5 others  
ORIGIN

Query Match 100.0%; Score 300; DB 13; Length 1071;  
Best Local Similarity 100.0%; Pred. No. 2.5e-72;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60  
|||||  
Db 359 ACGAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 418  
|||||

QY 61 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGGATACTTGACCTTGA 120  
|||||  
Db 419 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGGATACTTGACCTTGA 478  
|||||

QY 121 TGGAGATCAACACACAGCGGGACTTTCTCCACACAGCGCTCAACACATGTACAACTCC 180  
|||||  
Db 479 TGGAGATCAACACACAGCGGGACTTTCTCCACACAGCGCTCAACACATGTACAACTCC 538  
|||||

QY 181 GCACGAACTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGTGT 240  
|||||  
Db 539 GCACGAACTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGTGT 598  
|||||

QY 241 GCAGGCGGCTTCTGCTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 300  
|||||  
Db 599 GCAGGCGGCTTCTGCTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 658  
|||||

VERSION AL560880.2 GI:31285009  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 15, 2001 this sequence version replaced gi:12907768.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
Bp 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7987.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DL005AF100P1&cluster=7987.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DL005AF100P1.  
Location/Qualifiers  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL005YK19"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
25-NORMALIZED"  
Note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 322 a 267 c 292 g 283 t 37 others  
ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 1201;  
Best Local Similarity 100.0%; Pred. No. 2.6e-72;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60  
|||||  
Db 426 ACGAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 485  
|||||

QY 61 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGGATACTTGACCTTGA 120  
|||||  
Db 486 CTGCTGACAGCTTTGTGACACAGCAGGAGGCACATGCCAAGCTGGGATACTTGACCTTGA 545  
|||||

QY 121 TGGAGATCAACACAGCGGGACTTTCTCCACACAGCGCTCAACACATGTACAACTCC 180  
|||||  
Db 546 TGGAGATCAACACAGCGGGACTTTCTCCACACAGCGCTCAACACATGTACAACTCC 605  
|||||

QY 181 GCACGAACTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGTGT 240  
|||||  
Db 606 GCACGAACTCCAGCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCGTGT 665  
|||||

QY 241 GCAGGCGGCTTCTGCTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 300  
|||||  
Db 666 GCAGGCGGCTTCTGCTGGGGATGTGAGCGCTCAGGAGCTGATGAGGTACTCGTGGTCTGG 725  
|||||

RESULT 15  
AL582250/c  
LOCUS  
DEFINITION AL582250 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
ACCESSION AL582250  
VERSION AL582250.2 GI:31320467  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 16, 2001 this sequence version replaced gi:12950048.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7987.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DL005CC06NP1&cluster=7987.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DL005CC06NP1.  
Location/Qualifiers  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL005YF11"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 291 a 238 c 281 g 293 t 98 others  
ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 1201;  
Best Local Similarity 100.0%; Pred. No. 2.6e-72;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ACGAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 60  
Db 746 ACGAATCCGGACCCCTGGTCAAGGATATGTGGACACTCGTATAGCCAAACTCCGAGTGT 687  
QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGATTAACCTTGACCTTGA 120  
Db 686 CTGCTGACAGCTTTGTGAGACAGCAGGACATGCCAAGCTGGATTAACCTTGACCTTGA 627  
QY 121 TGGAGATCAACACACCGGGGACTTTCTCTACACAAAGCGCTCAACCACTGTACAAACTCC 180  
Db 626 TGGAGATCAACACACCGGGGACTTTCTCTACACAAAGCGCTCAACCACTGTACAAACTCC 567  
QY 181 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGTG 240  
Db 566 GCACGAACCTCCAGCCTCTGGAGAGTACTCAGTCTCAGGACTTCTAGAGAAAGGCTGTG 507  
QY 241 GCAGGCGGTGTGCTGGGGGATGTGAGCGCTCAGGAGTGTGAGTACTCGTGGTCTG 300  
Db 506 GCAGGCGGTGTGCTGGGGGATGTGAGCGCTCAGGAGTGTGAGTACTCGTGGTCTG 447

Search completed: October 14, 2003, 14:09:03  
Job time : 1279.67 secs

**THIS PAGE BLANK (USPTO)**

JOURNAL Patent: JP 2002191363-A 10946 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 COMMENT OS Homo sapiens (human)  
 PN JP 2002191363-A/10946  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC  
 10, C12P21/02, C12P1/68, C06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH key  
 Location/Qualifiers  
 FT CDS (37)..  
 TURES Location/Qualifiers  
 source 1..858  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 227 a 214 c 222 g 195 t  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 6; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-69;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACAGAAATCCGGACCCCTGCTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 DB 368 ACAGAAATCCGGACCCCTGCTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 427  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCGAAGCTGCAACCTTGACCTTGA 120  
 DB 428 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCGAAGCTGCAACCTTGACCTTGA 487  
 QY 121 TGGAGATCAACACAGCGGGACTTTCTCTACACAGCGCTCAACCATGTTACAACTCC 180  
 DB 488 TGGAGATCAACACAGCGGGACTTTCTCTACACAGCGCTCAACCATGTTACAACTCC 547  
 QY 181 GCAGGAACCTCCAGCCCTCTGAGAGTACTCAGTCTCAGGACTTCTACAGAAAGCCCTGGT 240  
 DB 548 GCAGGAACCTCCAGCCCTCTGAGAGTACTCAGTCTCAGGACTTCTACAGAAAGCCCTGGT 607  
 241 GCAGGCGGCTTGTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 300  
 608 GCAGGCGGCTTGTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 667  
 RESULT 2  
 AK001275 858 bp mRNA linear PRI 01-AUG-2002  
 LOCUS Homo sapiens cDNA FLJ10413 fis, clone NT2RP1000063.  
 DEFINITION  
 ACCESSION AK001275  
 VERSION AK001275.1 GI:7022428  
 KEYWORDS oligo capping; fls (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 REFERENCE 2 (bases 1 to 858)  
 AUTHORS Isogai, T. and Otsuki, T.  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Direct Submission  
 Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
 Location/Qualifiers  
 source 1..858  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP1000063"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP1"  
 /note="cloning vector: pUC19FL3-mRNA from NT2 neuronal precursor cells after 48-hours retinoic acid (RA) induction."  
 CDS 37..594  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAA91595.1"  
 /db\_xref="GI:7022429"  
 /translation="MDAAEVEFLAEKELVLIIPNFSLDKLYLIGDGLPFPNGLPVEV PLWAI NLKOROKRLLPPPEMDVEXLEKWRDHERKEETFTPMPSPIYMEITKLLNH ASDNIPKADEIRTLVDMNDTRIAKLVSADSFVRQOEHAHAKLDNLMLBINTSGTFL TQALNHYKRLTNLOPUESTQSQDF"  
 BASE COUNT 227 a 214 c 222 g 195 t  
 ORIGIN  
 Query Match 100.0%; Score 300; DB 9; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-69;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACAGAAATCCGGACCCCTGCTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 60  
 DB 368 ACAGAAATCCGGACCCCTGCTCAAGATATGTGGGACACTCGTATAGCCAAACTCCGAGTGT 427  
 QY 61 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCGAAGCTGCAACCTTGACCTTGA 120  
 DB 428 CTGCTGACAGCTTTGTGAGACAGCAGGAGGCACATGCGAAGCTGCAACCTTGACCTTGA 487  
 QY 121 TGGAGATCAACACAGCGGGACTTTCTCTACACAGCGCTCAACCATGTTACAACTCC 180  
 DB 488 TGGAGATCAACACAGCGGGACTTTCTCTACACAGCGCTCAACCATGTTACAACTCC 547  
 QY 181 GCAGGAACCTCCAGCCCTCTGAGAGTACTCAGTCTCAGGACTTCTACAGAAAGCCCTGGT 240  
 DB 548 GCAGGAACCTCCAGCCCTCTGAGAGTACTCAGTCTCAGGACTTCTACAGAAAGCCCTGGT 607  
 241 GCAGGCGGCTTGTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 300  
 608 GCAGGCGGCTTGTGGGGATGTGAGCGCTCAGGACGTGATGAGTACTCGTGGTCTGG 667  
 RESULT 3  
 AF151880 1174 bp mRNA linear PRI 18-MAY-2000  
 LOCUS Homo sapiens CGI-122 protein mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF151880  
 VERSION AF151880.1 GI:4929712  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1174)





**TITLE** Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics

**JOURNAL** Genome Res. 10 (5), 703-713 (2000)

**MEDLINE** 20272150

**PUBMED** 10810093

**REFERENCE** 2. (bases 1 to 1174)

**AUTHORS** Lin, W.-C.

**TITLE** Direct Submission

**JOURNAL** Sinica, No. 128, Sec. II, Academia Road, Taipei 115, Taiwan

**FEATURES** Location/Qualifiers

source 1..1174

1. /organism="Homo sapiens"

2. /mol\_type="mRNA"

3. /db\_xref="taxon:9606"

4. /chromosome="16"

5. 77..634

6. /codon\_start=1

7. /product="CGI-122 protein"

8. /protein\_id="A034117.1"

9. /db\_xref="GI:4929713"

10. /translation="MDAAEVEFLAEKELVTIIPNFDKIYLGDDLPNPLPVEV PLWAIILKOROKRLLPPEDMDVEKLEMRDHERKEETFTPMSPYMETLKLLNH ASDNIPKADEIRTLVKMDWTRIAKLRSADSFVRQEAHAKLDNLTLMEINTSGTFL TOALNHMYKLRNLQPLESTOSQDF"

BASE COUNT 300 a 299 c 307 g 268 t

ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 1174;

Best Local Similarity 100.0%; Pred. No. 4.5e-69;

Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGACCCCTGCTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60

DB 408 ACGAAATCCGACCCCTGCTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 467

QY 61 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCCAAGCTGGATACCTTGACTTGA 120

DB 468 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCCAAGCTGGATACCTTGACTTGA 527

QY 121 TGGAGATCAACACCGCGGACTTTCTCTCACAAAGCGCTCAACACATGTACAACTCC 180

DB 528 TGGAGATCAACACCGCGGACTTTCTCTCACAAAGCGCTCAACACATGTACAACTCC 587

QY 181 GCACGAACCTCCAGCCTCTGGAGACTCTCAGTCTCAGGACTCTTACAGAAAGGCTGT 240

DB 588 GCACGAACCTCCAGCCTCTGGAGACTCTCAGTCTCAGGACTCTTACAGAAAGGCTGT 547

QY 241 GCAGCGGCTTGTGGGGATGTAGCGCTCAGGAGTGTAGGAGTGTAGGAGTGTAGG 300

DB 648 GCAGCGGCTTGTGGGGATGTAGCGCTCAGGAGTGTAGGAGTGTAGGAGTGTAGG 707

**RESULT 4**

**LOCUS** BC010164 1180 bp mRNA linear PRI 12-JUL-2001

**DEFINITION** Homo sapiens, HSPC037 protein, clone MGC:19836 IMAGE:4098007, mRNA, complete cds.

**ACCESSION** BC010164

**VERSION** BC010164.1 GI:14603431

**KEYWORDS** MGC.

**SOURCE** Homo sapiens (human)

**ORGANISM** Homo sapiens

**REFERENCE** 1. (bases 1 to 1180)

**AUTHORS** Strausberg, R.

**TITLE** Direct Submission

**JOURNAL** Submitted (02-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK** NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
nisc\_mgc@hgr.nih.gov  
Contact: Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 27 Row: 1 Column: 5  
This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction, similarity but not identity to protein.

FEATURES

Location/Qualifiers

1..1180

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:19836 IMAGE:4098007"

/tissue\_type="Brain, neuroblastoma"

/clone\_lib="NIH\_MGC\_19"

/lab\_host="DH10B-R"

/note="Vector: pONB7"

68..625

/codon\_start=1

/product="HSPC037 protein"

/protein\_id="AAH10164.1"

/db\_xref="GI:14603431"

/translation="MDAAEVEFLAEKELVTIIPNFDKIYLGDDLPNPLPVEV PLWAIILKOROKRLLPPEDMDVEKLEMRDHERKEETFTPMSPYMETLKLLNH ASDNIPKADEIRTLVKMDWTRIAKLRSADSFVRQEAHAKLDNLTLMEINTSGTFL TOALNHMYKLRNLQPLESTOSQDF"

BASE COUNT 316 a 292 c 307 g 265 t

ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 1180;

Best Local Similarity 100.0%; Pred. No. 4.5e-69;

Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAAATCCGACCCCTGCTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 60

DB 399 ACGAAATCCGACCCCTGCTCAAGGATATGTGGACACTCGTATAGCCAACTCCGAGTGT 458

QY 61 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCCAAGCTGGATACCTTGACTTGA 120

DB 459 CTGCTGACAGCTTTGTGAGACAGGAGGACATGCCAAGCTGGATACCTTGACTTGA 518

QY 121 TGGAGATCAACACCGCGGACTTTCTCTCACAAAGCGCTCAACACATGTACAACTCC 180

DB 519 TGGAGATCAACACCGCGGACTTTCTCTCACAAAGCGCTCAACACATGTACAACTCC 578

QY 181 GCACGAACCTCCAGCCTCTGGAGACTCTCAGTCTCAGGAGTGTAGGAGTGTAGG 240

DB 579 GCACGAACCTCCAGCCTCTGGAGACTCTCAGTCTCAGGAGTGTAGGAGTGTAGG 638

QY 241 GCAGCGGCTTGTGGGGATGTAGCGCTCAGGAGTGTAGGAGTGTAGGAGTGTAGG 300

DB 639 GCAGCGGCTTGTGGGGATGTAGCGCTCAGGAGTGTAGGAGTGTAGGAGTGTAGG 698

**RESULT 5**

THIS PAGE BLANK (USPTO)

8-17-2